

ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-547-267-3

Query Match 44.7%; Score 686.5; DB 4; Length 303;
Best Local Similarity 48.7%; Pred. No. 1.6e-67;
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;
1 MAVGSKSPATSTLFDKTRSVLMYAWCHCDVDDQTLGFHADQPSQMPQRLQQ 60
12 IAQGSQFAAKLMPGPIREDTVMYAWCHHADVDIGQVMSAPEAGGD--PQALGA 69
61 LEMKTRQAYAGSOMEH-----PAPAAFOEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTL 115
70 LEADTLAA-----LHEDGPMSPPPAALQVARRHDFPDLMPMDLIEGFAMDVADREYRSL 124
116 DDTLYCYHVAGVGLMMAQIMGVDRDNATLDRACDLGLAFOLITNIARDIVDDAQVGRCYL 175
125 DDVLEYSYHVAGVGVMMARVMGVQDDAVLDRACDLGLAFOLITNIARDIVDDAAIGRCYL 184
176 PESWLEERGLTKANYAAPENQALSRIRAGRLVRAEPYVYSSMAGLAQLPLRSAAWATA 235
185 PADWLAEG---ATVEGVPSDALYSVIIRLDRAEPYASARQGLPHLPFRCAWSIAAA 241
236 KQVYRKIGVKVEQAGKQAWDRHQSTSTAELTLLLTASGQAVTSMKTYPPRPAHLWQRP 295
242 LRIYRAIGTRIRQGGPEAYRQRISTSKAKIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

RESULT 13
3-09-920-923B-3
Sequence 3, Application US/09920923B
Patent No. 6671134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/125944)
CURRENT APPLICATION NUMBER: US/09/920-923B
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent in version 3.1
SEQ ID NO 3
LENGTH: 303
TYPE: PRP
ORGANISM: Flavobacterium sp. R1534

US-09-920-923B-3

Query Match 44.7%; Score 686.5; DB 4; Length 303;
Best Local Similarity 48.7%; Pred. No. 1.6e-67;
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;
1 MAVGSKSPATSTLFDKTRSVLMYAWCHCDVDDQTLGFHADQPSQMPQRLQQ 60
12 IAQGSQFAAKLMPGPIREDTVMYAWCHHADVDIGQVMSAPEAGGD--PQALGA 69
61 LEMKTRQAYAGSOMEH-----PAPAAFOEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTL 115
70 LEADTLAA-----LHEDGPMSPPPAALQVARRHDFPDLMPMDLIEGFAMDVADREYRSL 124
116 DDTLYCYHVAGVGLMMAQIMGVDRDNATLDRACDLGLAFOLITNIARDIVDDAQVGRCYL 175
125 DDVLEYSYHVAGVGVMMARVMGVQDDAVLDRACDLGLAFOLITNIARDIVDDAAIGRCYL 184
176 PESWLEERGLTKANYAAPENQALSRIRAGRLVRAEPYVYSSMAGLAQLPLRSAAWATA 235
185 PADWLAEG---ATVEGVPSDALYSVIIRLDRAEPYASARQGLPHLPFRCAWSIAAA 241
236 KQVYRKIGVKVEQAGKQAWDRHQSTSTAELTLLLTASGQAVTSMKTYPPRPAHLWQRP 295
242 LRIYRAIGTRIRQGGPEAYRQRISTSKAKIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

RESULT 14
US-08-579-667-6
Sequence 6, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: No. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-579-667-6

Query Match 20.7%; Score 318.5; DB 1; Length 410;
Best Local Similarity 28.6%; Pred. No. 1.3e-26;

130	Db	AKTFYLTGTLMTTPERRAINAIWYCRRTDELVDGPN-----ASHITPQGLDRWSDL	181
65	QY	TRQAYAGSQGHEPAPAFQEVAMAH--DIAPAYAFDHLGFMADVSTRYLTLDLTRYC	122
182	Db	LEDFVSG-RPFMDLDAALSDTVSKFPVDIQPFR--DMIEGRMDLRKSRVNFDELVLVC	238
123	QY	YHVAGVGLMAAQINGVRDIA-----TLDBACDLGLAFOLTNIARDIVDDAQVGRCYLP	176
239	Db	YVYVAGTVGLMSVPIINGIAPDSKATTEVYVYNAALGALIANQLTILRDVGDGDAARRGVYLP	298
177	QY	ESWLEEBEGLTANTAAAPENRQALSKRIAGLVRREAPYVYVSSMAGLAQLPUSAWAATAK	236
299	Db	QDELAHAGLSDDDIIFAGKVTDKYSFPMKKIQIQRARKEFDEAREGVQLSSASRPVWASL	358
237	QY	QYVRKIGYKVEQAGQAWDHRQSTSTAEKLTLLLTASGQAVTGRMKT	283
359	Db	LIYVOILDEIRANDYNNSTKRAYVYSKSKLIISLPIANAKSLVPTPT	405

RESULT 15
 S-08-579-667-2
 Sequence 2, Application US/08579667
 Patent NO. 5705624
 GENERAL INFORMATION:
 APPLICANT: Fitzmaurice, Wayne P.
 APPLICANT: Hellmann, Gary M.
 APPLICANT: Grill, Laurence K.
 APPLICANT: Kumagai, Monto H.
 APPLICANT: Della-Cioppa, Guv R.
 TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
 TITLE OF INVENTION: PHYTOENE BIOSYNTHESIS
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Virginia C. Bennett
 STREET: 1211 East Morehead Street, PO Drawer 34009
 CITY: Charlotte
 STATE: No. 5705624th Carolina
 COUNTRY: USA
 ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-579-667-2

```
Query Match      20.6%   Score 315.5; DB 1; Length 410;
Best Local Similarity 28.6%; Pred. NO. 2.9e-26;
Matches 82; Conservative 57; Mismatches 129; Indels 19; Gaps 5;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:26:38 ; Search time 34.9969 Seconds
(without alignments)
2389.754 Million cell updates/sec

tle: US-09-941-947A-34

fect score: 1535
quence: 1 MAVGSKSFATSTLFDKTR.....VTRSMKTYPPRPAPHLWQRP 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

al number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A Geneseq 29Jan04.*

- 1: geneseq1980s.*
- 2: geneseq1990s.*
- 3: geneseq2000s.*
- 4: geneseq2001s.*
- 5: geneseq2002s.*
- 6: geneseq2003as.*
- 7: geneseq2003bs.*
- 8: geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1535	100.0	296	5	AEE22315 Pantoea s
2	1535	100.0	296	6	AAO16022 Pantoea s
3	1535	100.0	296	6	ABP96689 Pantoea s
4	1368	89.1	296	2	AAW87887 Protein e
5	1368	89.1	296	2	AAW99097 Erwinia u
6	1368	89.1	309	2	AAW93845 E. uredo
7	1363	88.8	296	2	AAW07467 Polypepti
8	1363	88.8	309	2	AAW82256 C. utilis
9	1028	67.0	309	2	AAW16492 Erwinia h
10	997.5	65.0	308	2	AAW13983 Phytoene
11	997.5	65.0	308	2	AAW01121 Phytoene
12	997.5	65.0	308	2	AAW00172 E. herbic
13	997.5	65.0	308	2	AAW00343 Phytoene
14	997.5	65.0	308	2	AAW32471 Erwinia h
15	904	58.9	309	6	ABM70124 Phototrab
16	686.5	44.7	303	2	AAW06516 Flavobact
17	686.5	44.7	303	2	AAW69531 Flavobact
18	336.5	21.9	440	5	AAW51841 Nicotiana
19	318.5	20.7	410	2	AAW41059 Phytoene
20	315.5	20.6	410	2	AAW41057 Phytoene
21	311.5	20.3	410	2	AAW41060 Phytoene
22	311.5	20.3	413	2	AAW41058 Phytoene
23	310.5	20.2	410	5	AAW51842 Nicotiana
24	309.5	20.2	400	3	AAW44217 Soybean p
25	307	20.0	363	3	AAG10660 Arabidops

ALIGNMENTS

RESULT 1
AAE22315
ID AAE22315 standard; protein; 296 AA.

XX AC AAE22315;

DT 25-JUL-2002 (first entry)

DE Pantoea stewartii phytoene synthase (CrtB) enzyme.

XX Pantoea stewartii.

KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KM anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; phytoene synthase; CrtB.

XX Pantoea stewartii.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-0229858P.

XX 01-SEP-2000; 2000US-0229907P.

XX (DUFO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odom JM, Picataggio SK, Rouviere PB;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35513.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates.

XX Claim 17; Page 144-145; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds. The method comprises a transformed metabolising host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the control of regulatory sequences, and contacting the host cell with carbon substrate to produce a carotenoid compound. The method is useful for producing carotenoid compounds such as antheraxanthin and astaxanthin, by using microorganism having a nucleic acid molecule encoding enzymes in

CC the carotenoid biosynthetic pathway and which metabolise single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids, flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Pantoea stewartii phytoene synthase (CrtB) enzyme used in the
 CC invention
 CC Sequence 296 AA;

Query Match 100.0%; Score 1535; DB 5; Length 296;
 Best Local Similarity 100.0%; Pred. No. 1.7e-167;
 Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVGSKSFATASTLFDKTRRSVLMYAMCRHCDVDVDDQTLGFHADQPSQMEQRLQQ 60
 1 MAVGSKSFATASTLFDKTRRSVLMYAMCRHCDVDVDDQTLGFHADQPSQMEQRLQQ 60
 61 LEMKTRQAYAGSQMHEPFAFAQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
 61 LEMKTRQAYAGSQMHEPFAFAQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
 121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
 241 KIGYKVEQAGKQANDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWORPI 296
 241 KIGYKVEQAGKQANDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWORPI 296

RESULT 2

VAO16022
 ID AAO16022 standard; protein; 296 AA.
 AC AAO16022;
 XT 20-FEB-2003 (first entry)
 SE Pantoea stewartii phytoene synthase.
 CW Carotenoid; crt.
 JS Pantoea stewartii.
 FN WO200279395-A2.
 PD 10-OCT-2002.
 PF 25-JAN-2002; 2002WO-US002124.
 PR 26-JAN-2001; 2001US-0264329P.
 PR 04-MAY-2001; 2001US-0288984P.
 PR (CRG) CARGILL INC.
 PI De Souza ML, Koilmann SR, May CA, Schroeder WB;
 DR WPI; 2003-075455/07.
 DR N-PSDB; ABL14194.

Novel isolated nucleic acid useful e.g. to engineer host cells with the
 PT ability to produce particular carotenoids and polypeptides useful in cell
 PT -free systems to make particular carotenoids.

XX Claim 32; Page 64-65; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of
 CC carotenoid (crt)-related proteins. The crt-related DNA and protein

CC sequences of the invention are useful for engineering cells which are
 CC able to produce carotenoids. The present amino acid sequence represents a
 CC crt-related protein of the invention
 XX Sequence 296 AA;

Query Match 100.0%; Score 1535; DB 6; Length 296;
 Best Local Similarity 100.0%; Pred. No. 1.7e-167;
 Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVGSKSFATASTLFDKTRRSVLMYAMCRHCDVDVDDQTLGFHADQPSQMEQRLQQ 60
 1 MAVGSKSFATASTLFDKTRRSVLMYAMCRHCDVDVDDQTLGFHADQPSQMEQRLQQ 60
 61 LEMKTRQAYAGSQMHEPFAFAQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
 61 LEMKTRQAYAGSQMHEPFAFAQEVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTDDTLR 120
 121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 121 YCYHVAGVVGGLMMAQIMGVRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQVYR 240
 241 KIGYKVEQAGKQANDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWORPI 296
 241 KIGYKVEQAGKQANDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWORPI 296

RESULT 3

ABP96689
 ID ABP96689 standard; protein; 296 AA.
 XX AC ABP96689;
 XX 03-JUN-2003 (first entry)
 XX Pantoea stewartii phytoene synthase SEQ ID NO:10.
 DE Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
 KW crtI; crtB; crtZ; phytoene synthase; enzyme; phytoene; carotenoid.
 XX Pantoea stewartii.
 OS Pantoea stewartii.
 FN WO2003016503-A2.
 PD 27-FEB-2003.
 XX 15-AUG-2002; 2002WO-US026647.
 XX 15-AUG-2001; 2001US-0312646P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;
 PI WPI; 2003-268323/26.
 DR N-PSDB; ACC44763.
 DR Novel nucleic acid molecule isolated from Pantoea stewartii encoding a
 PT carotenoid biosynthetic enzyme, useful for regulating carotenoid
 PT biosynthesis in an organism.
 XX Claim 4; Page 66; 68pp; English.

XX The present invention describes Pantoea stewartii carotenoid biosynthetic
 CC enzymes (I). More specifically described are the geranylgeranyl
 CC pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
 CC lycopene cyclase (crtI), phytoene desaturase (crtI), phytoene synthase
 CC (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to
 CC ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating

carotenoid biosynthesis in an organism, by over-expressing (I) in an organism, such that the carotenoid biosynthesis is altered in the organism. (I) and the genes encoding (I) are useful for converting phytoene to the carotenoids, for creating recombinant organisms that have the ability to produce various carotenoid compounds, and also for enhancing or manipulating carotenoid compounds. (I) can also be used for producing gene products having enhanced or altered activity

Sequence 296 AA;

Query Match 100.0%; Score 1535; DB 6; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.7e-167; Mismatches 0; Indels 0; Gaps 0;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVGSKSPATATLFDKTRRSVLMVYAWCRHCDVDDQTLGFHADQPSQMPERLQQ 60
1 MAVGSKSPATATLFDKTRRSVLMVYAWCRHCDVDDQTLGFHADQPSQMPERLQQ 60
61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120
61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120
121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240
241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296

RESULT 4

AAW87887
AAW87887 standard; protein; 296 AA.
AAW87887;
17-OCT-2003 (revised)
10-MAY-1999 (first entry)
Protein encoded by the carotenoid biosynthesis gene crtB.
Carotenoid biosynthesis; astaxanthin diglucoside; crtB gene; crtB gene;
crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;
carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;
food additive.
Pantoea ananatis.
JP10327865-A.
15-DEC-1998.
29-MAY-1997; 97JP-00140460.
29-MAY-1997; 97JP-00140460.
(KIRI) KIRIN BREWERY KK.
(KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
WPI; 1999-099030/09.
N-PSDB; AAV84079.
New carotenoid glucoside(s) - used as food additives.

Disclosure; Page 15-16; 26pp; Japanese.

The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and

adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid glycoside, in which all, or part of, carotenoid biosynthesis genes crtB, crtB, crtI, crtY, crtZ, crtX or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 296 AA;

Query Match 89.1%; Score 1368; DB 2; Length 296;
Best Local Similarity 88.9%; Pred. No. 2.8e-148; Mismatches 11; Indels 0; Gaps 0;
Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

1 MAVGSKSPATATLFDKTRRSVLMVYAWCRHCDVDDQTLGFHADQPSQMPERLQQ 60
1 MAVGSKSPATATLFDKTRRSVLMVYAWCRHCDVDDQTLGFHADQPSQMPERLQQ 60
61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120
61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120
121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
121 YCYHVAGVVGGLMMAQIMGVNRDNATLDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSWAGLAQLPLRSAWAIAATAKQVYR 240
241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
241 KIGVVEQAGQAMDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296

RESULT 5

AAW99097
ID AAW99097 standard; protein; 296 AA.
XX AAW99097;
AC AAW99097;
17-OCT-2003 (revised)
14-MAY-1999 (first entry)
Erwinia uredovora crtB protein sequence.
XX Beta-carotene hydroxylase; crtY; crtB; crtE; crtI; xanthophyll;
XX metabolite.
XX Pantoea ananatis.
XX JP11046770-A.
23-FEB-1999.
07-AUG-1997; 97JP-00213648.
07-AUG-1997; 97JP-00213648.
(KIRI) KIRIN BREWERY KK.
WPI; 1999-208113/18.
N-PSDB; AAX19117.
Beta-carotene hydroxylase - useful for preparation of xanthophylls and their metabolites.
Disclosure; Page 11-12; 17pp; Japanese.

The present invention describes beta-carotene hydroxylase. Beta-carotene can be used in the preparation of xanthophylls and their metabolites. The present sequence represents an Erwinia uredovora crtB protein sequence

Q Sequence 296 AA;

Query Match 88.8%; Score 1363; DB 2; Length 296;
 Best Local Similarity 88.5%; Pred. No. 1e-147;
 Matches 262; Conservative 11; Mismatches 23; Indels 0; Gaps 0;

1 MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDIDDQTLGFHADOPSSQMPERLQQ 60
 1 MAVGSKSFATASKLFDKTRRSVLMYAWCRHCDVDIDDQTLGFQARQALQTPQRLMQ 73
 61 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 120
 61 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 120
 121 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 121 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 121 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 181 EBEGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAWAIAATAKQVYR 240
 181 EBEGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAWAIAATAKQVYR 240
 241 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296
 241 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296

RESULT 8

AW82256
 AAW82256 standard; protein; 309 AA.
 AAW82256;
 17-OCT-2003 (revised)
 16-JUL-1999 (first entry)
 C. utilis crtB protein.
 HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtB;
 carotenoid.
 Pichia jadinii.
 JP10248575-A.
 22-SEP-1998.
 12-MAR-1997; 97JP-00058012.
 12-MAR-1997; 97JP-00058012.
 (KIRI) KIRIN BREWERY KK.
 WPI; 1998-560727/48.
 N-PSDB; AAW73180.

Gene useful for increase in carotenoid production - and preparation of carotenoid.

Example 2; Fig 8-10; 54pp; Japanese.

This invention describes a novel method for the preparation of carotenoids using genes and proteins isolated from *Candida utilis*. The invention specifically describes the isolation of a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence represents the *Candida utilis* crtB protein which is used in the method of the invention. (Updated on 17-OCT-2003 to standardise OS field)

Q Sequence 309 AA;

Query Match 88.8%; Score 1363; DB 2; Length 309;
 Best Local Similarity 88.5%; Pred. No. 1.1e-147;
 Matches 262; Conservative 11; Mismatches 23; Indels 0; Gaps 0;

QY 1 MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDIDDQTLGFHADOPSSQMPERLQQ 60
 DB 14 MAVGSKSFATASKLFDKTRRSVLMYAWCRHCDVDIDDQTLGFQARQALQTPQRLMQ 73
 QY 61 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 120
 DB 74 LEMKTRQAVAGSOMHEPAPAFQEVAMAHDIAPAFDHLGFGFAMDVRETRYLTLDLTLR 133
 QY 121 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 DB 134 YCVHVAGVVGGLMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193
 QY 181 EBEGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAWAIAATAKQVYR 240
 DB 194 EBEGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAWAIAATAKQVYR 253
 QY 241 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296
 DB 254 KIGVKVEQAGQAWDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 309

RESULT 9

AAW16492
 ID AAW16492 standard; protein; 309 AA.
 AC AAW16492;
 XX
 DT 17-OCT-2003 (revised)
 DT 22-JUL-1997 (first entry)
 XX
 DE Brwinia herbicola phytoene synthase.
 XX
 KW Transgenic plant; selectable marker; carotenoid; pigment;
 KW phytoene synthase; lycopene; plasmid pET0203.
 XX
 OS Pantoea agglomerans.
 XX
 PN WO9714807-A1.
 XX
 PD 24-APR-1997.
 XX
 PP 29-MAR-1996; 96WO-US004313.
 XX
 PR 16-OCT-1995; 95US-00543608.
 XX
 PA (SEMI-) SEMINIS VEGTABLES.
 XX
 PI Trulson AJ, Braun CU;
 XX
 DR WPI; 1997-245122/22.
 DR N-PSDB; AAT66534.
 XX
 XX

Visual identification of transgenic plant material - from production of carotenoid pigment encoded by cassette containing Brwinia phytoene synthase gene, useful for selecting material for regeneration.

Example; Page 33-34; 62pp; English.

Brwinia herbicola phytoene synthase (AAW16492) catalyses the biosynthesis of phytoene, a precursor of the red carotenoid lycopene, from geranylgeranyl pyrophosphate. Its expression in transgenic plant material can be utilised as a means of visually distinguishing such material from non-transgenic cells and tissues. Novel binary vector pET0203 (ATCC 97282) includes an expression cassette comprising the tomato E8 promoter (see also AAT66533), a plastid targeting signal (AAT66535) fused to the phytoene synthase gene (AAT66534), and a 3' non-translated region (AAT66536). Transgenic plant cells and tissues are identified by the appearance of orange colour. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 309 AA;

Query Match 67.0%; Score 1028; DB 2; Length 309;
 Best Local Similarity 64.9%; Pred. No. 4e-109;
 Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

2Y 1 MAVSKSFATASTILFDKATRSVLMLYAWCHDDVDQTLGFHADQPSQMPQRLQ 60
 14 MANGSKSFATAAKLFDPATRSVLMLYTCRCHDDVDDQTHGFASAAAEEATQRLAR 73

2Y 61 LEMKTRQAYAGSQMHPEPAFAAFQEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTLDLTLR 120
 74 LRTLTAAFEAGAQMDPAFAAFQEVVALTHGITPRMALDHLDFAMDVQATRYVTFEDTLR 133

2Y 121 YCYHVAGVGLMAQINGVDNATLDRACDLGLAFQLTNTIARDIVDDAQVGRCYLPESWL 180
 134 YCYHVAGVGLMARVNGVDRVLDRACDLGLAFQLTNTIARDIIDDADIDRCYLPESWL 193

2Y 181 BEEGLTKANYAAPENRQALSRIGRLVREAEPPYVVSMAGLQPLRSAWAIATAKQVYR 240
 194 QDAGLTPENYAARENRPALARWR-RUJIDAAEPYVVSQAGLHDLPRCAWAIATARSVYR 253

2Y 241 KIGYVFCAGQAWDHQSTSTAKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
 254 EIGIKVKAAGSAMDWRQHTSKGEKIAMLAAPQGVIRAKTRTTPRPAGLWQRPV 309

RESULT 10
 AAW01121 standard; protein; 308 AA.

XX AAW01121;
 XX AC AAW01121;
 XX DT 16-OCT-2003 (revised)
 XX DT 25-MAR-2003 (revised)
 XX DT 11-DEC-1996 (first entry)
 XX XX
 XX DE Phytoene synthase.
 XX KW GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
 XX KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
 XX KW pigment; food colourant; chloroplast transit peptide; increase yield;
 XX KW tobacco ribulose bis-phosphate carboxylase-oxygenase.
 XX OS Pantoea agglomerans.
 XX PN US5530188-A.
 XX PD 25-JUN-1996.
 XX XX
 XX PF 21-JUL-1993; 93US-00095726.
 XX PR 02-MAR-1990; 90US-00487613.
 XX PR 18-MAY-1990; 90US-00525551.
 XX PR 03-AUG-1990; 90US-00562674.
 XX PR 28-FEB-1991; 91US-00662921.
 XX PR 30-OCT-1991; 91US-00785566.
 XX XX
 XX PA (STAD) AMOCO CORP.
 XX PI Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
 XX PI Ausich RL;
 XX XX
 XX DR WPI; 1996-308823/31.
 XX DR N-PSDB; AAT40791.
 XX XX

CC AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726
 CC respectively. Recombinant expression plasmids can be used to produce
 CC large amts. of the enzymes and hence large amts. of the carotenoids which
 CC they synthesize. Carotenoids are pigments with a variety of applications.
 CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-2003 to
 CC correct PR field.) (Updated on 24-OCT-2003 to standardise OS field)
 XX Sequence 308 AA;

Query Match 65.0%; Score 997.5; DB 2; Length 308;
 Best Local Similarity 63.9%; Pred. No. 1.3e-105;
 Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

QY 1 MAVSKSFATASTILFDKATRSVLMLYAWCHDDVDQTLGFHADQPSQMPQRLQ 60
 14 MANGSKSFATAAKLFDPATRSVLMLYTCRCHDDVDDQTHGFASAAAEEATQRLAR 73

QY 61 LEMKTRQAYAGSQMHPEPAFAAFQEVAMAHDIAPAYAFDHLGPFAMDVRETRYLTLDLTLR 120
 74 LRTLTAAFEAGAQMDPAFAAFQEVVALTHGITPRMALDHLDFAMDVQATRYVTFEDTLR 133

QY 121 YCYHVAGVGLMAQINGVDNATLDRACDLGLAFQLTNTIARDIVDDAQVGRCYLPESWL 180
 134 YCYHVAGVGLMARVNGVDRVLDRACDLGLAFQLTNTIARDIIDDADIDRCYLPESWL 193

QY 181 BEEGLTKANYAAPENRQALSRIGRLVREAEPPYVVSMAGLQPLRSAWAIATAKQVYR 240
 194 QDAGLTPENYAARENRPALARWR-RUJIDAAEPYVVSQAGLHDLRRSAWAIATARSVYR 252

QY 241 KIGYVFCAGQAWDHQSTSTAKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 296
 253 EIGIKVKAAGSAMDWRQHTSKGEKIAMLAAPQGVIRAKTRTTPRPAGLWQRPV 308

RESULT 11
 AAW01121
 ID AAW01121 standard; protein; 308 AA.

XX AC AAW01121;
 XX DT 16-OCT-2003 (revised)
 XX DT 25-MAR-2003 (revised)
 XX DT 11-DEC-1996 (first entry)
 XX XX
 XX DE Phytoene synthase.
 XX KW GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
 XX KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
 XX KW pigment; food colourant; chloroplast transit peptide; increase yield;
 XX KW tobacco ribulose bis-phosphate carboxylase-oxygenase.
 XX OS Pantoea agglomerans.
 XX PN US5530188-A.
 XX PD 25-JUN-1996.
 XX XX
 XX PF 21-JUL-1993; 93US-00095726.
 XX PR 02-MAR-1990; 90US-00487613.
 XX PR 18-MAY-1990; 90US-00525551.
 XX PR 03-AUG-1990; 90US-00562674.
 XX PR 28-FEB-1991; 91US-00662921.
 XX PR 30-OCT-1991; 91US-00785566.
 XX XX
 XX PA (STAD) AMOCO CORP.
 XX PI Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
 XX PI Ausich RL;
 XX XX
 XX DR WPI; 1996-308823/31.
 XX DR N-PSDB; AAT40791.
 XX XX

Increasing prodn. of total carotenoid(s) in a higher plant - by transforming with vector encoding chloroplast transit peptide operably linked to the Erwinia herbicola lycopene cyclase structural gene.

Example 4; Fig 4; 99pp; English.

The present sequence is that of the Erwinia herbicola phytoene synthase which acts on geranylgeranyl pyrophosphate (GGPP) (produced by the action of GGPP synthase (AAW0119), on isopentenyl pyrophosphate and dimethylallyl pyrophosphate). GGPP is a 20-carbon atom precursor of phytoene, the first carotenoid in the carotenoid biosynthesis pathway. Production of total carotenoids in a plant can be increased by transforming the plant with DNA encoding enzymes involved in the biosynthesis pathway, in particular the lycopene cyclase gene (AAT40795). Lycopene is prepared biosynthetically from phytoene through four sequential dehydrogenation reactions which can be carried out by a single dehydrogenase (AAT40793) in Erwinia sp. Beta-carotene is produced by the action of lycopene cyclase on lycopene. A chloroplast transit peptide of the tobacco ribulose bis-phosphate carboxylase-oxygenase gene (see AAT40794) is operatively linked in frame to the 5' end of the lycopene cyclase structural gene. This leads to increased production of total carotenoids in the chloroplast of transformed plants as compared to native, non-transformed plants of the same type. Beta-carotene is an effective and apparently harmless food colourant and is also in the pathway for biological synthesis of further C40 carotenoids such as zeaxanthin and zeaxanthin diglucoside. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 308 AA;

Query Match 65.0%; Score 997.5; DB 2; Length 308;

Best Local Similarity 63.9%; Pred. No. 1.3e-105;

Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

1 MAVGSKSPATASTLFDKATRSVLMLYAWCHCDVDVDDQTLGPHADQPSQMPQRLQQ 60

14 MANGSKSPATAKLFDPATRSVLMLYAWCHCDVDVDDQTLGPHADQPSQMPQRLQQ 73

61 LEMKTRQYAGSOMHEPAPAFQEVAMADHAPAYAFDHLGEGFAMDVRETRYLTLDPTLR 120

74 LRTLTAAFEAGQMDPAPAFQEVAMADHAPAYAFDHLGEGFAMDVRETRYLTLDPTLR 133

121 YCYHVAGVGLMMAQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPESWL 180

134 YCYHVAGVGLMMAQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPESWL 193

181 EEEGLTKANYAAPENRQALSRVREAPFYVSSWAGLAQPLPSAWAIATAKQVYR 240

194 QDAGLAPENYAARENRPALARWR-RLIDAAEPYVSSWAGLAQPLPSAWAIATAKQVYR 252

241 KIGVKVEAGQKQMDHRSQSTABKLTLLTASQAVTSRMKTYPPRPAHLWQRP 296

253 EIGIKVKAAGSAMDWRQHTSKGKIAMLAAPQGVIRAKTRVTRTPRPAHLWQRP 308

SULT 12

AW00172

AAW00172 standard; protein; 308 AA.

AAW00172;

16-OCT-2003 (revised)

25-MAR-2003 (revised)

17-OCT-1996 (first entry)

E. herbicola phytoene synthase encoded by parC285.

E. herbicola; geranylgeranyl pyrophosphate synthase; parC376;

GGPP synthase; biosynthesis; carotenoid; lycopene;

farnesyl pyrophosphate; phytoene; fpp; isopentenyl pyrophosphate; IPP;

tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;

food colourant; herbicide; norflurazon.

Pantoea agglomerans.

US5530189-A.

XX

25-JUN-1996.

XX

22-JUL-1993; 93US-00096043.

XX

02-MAR-1990; 90US-00487613.

PR

18-MAY-1990; 90US-00525551.

PR

03-AUG-1990; 90US-00562674.

PR

28-FEB-1991; 91US-00662921.

PR

30-OCT-1991; 91US-00785568.

XX

(STAD) AMOCO CORP.

XX

Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;

PI Ausich RL;

XX

WPI; 1996-308824/31.

DR

N-PSDB; AAT37093.

DR

DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for

PT prodn. of lycopene, and to produce transgenic plants resistant to

PT norflurazon.

XX

Example 4; Fig 4; 87pp; English.

XX

This sequence represents Erwinia herbicola phytoene synthase encoded by

CC plasmid pARC285. Phytoene synthase is an enzyme which is involved in the

CC biosynthesis of carotenoids, esp. lycopene, from the ubiquitous

CC precursor, farnesyl pyrophosphate. In E. herbicola, phytoene has been

CC found to be formed biosynthetically in a two-step process. The initial

CC step is the condensation of farnesyl pyrophosphate (fpp) and isopentenyl

CC pyrophosphate (IPP) to form GGPP. This reaction is catalysed by GGPP

CC synthase. This first step is immediately followed by a tail to tail

CC dimerisation of GGPP, catalysed by the enzyme phytoene synthase, to form

CC phytoene. Lycopene is produced from phytoene by the catalytic action of

CC phytoene dehydrogenase-4H. The genes encoding components of the lycopene

CC biosynthesis pathway may be used to transform a host cell for the

CC commercial production of lycopene which is used as a food colourant.

CC Plants transformed with the phytoene dehydrogenase-4H coding sequence are

CC protected from the herbicide norflurazon. (Updated on 25-MAR-2003 to

CC correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

XX

Sequence 308 AA;

Query Match 65.0%; Score 997.5; DB 2; Length 308;

Best Local Similarity 63.9%; Pred. No. 1.3e-105;

Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;

QY 1 MAVGSKSPATASTLFDKATRSVLMLYAWCHCDVDVDDQTLGPHADQPSQMPQRLQQ 60

Db 14 MANGSKSPATAKLFDPATRSVLMLYAWCHCDVDVDDQTLGPHADQPSQMPQRLQQ 73

QY 61 LEMKTRQYAGSOMHEPAPAFQEVAMADHAPAYAFDHLGEGFAMDVRETRYLTLDPTLR 120

Db 74 LRTLTAAFEAGQMDPAPAFQEVAMADHAPAYAFDHLGEGFAMDVRETRYLTLDPTLR 133

QY 121 YCYHVAGVGLMMAQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPESWL 180

Db 134 YCYHVAGVGLMMAQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCVLPESWL 193

QY 181 EEEGLTKANYAAPENRQALSRVREAPFYVSSWAGLAQPLPSAWAIATAKQVYR 240

Db 194 QDAGLAPENYAARENRPALARWR-RLIDAAEPYVSSWAGLAQPLPSAWAIATAKQVYR 252

QY 241 KIGVKVEAGQKQMDHRSQSTABKLTLLTASQAVTSRMKTYPPRPAHLWQRP 296

Db 253 EIGIKVKAAGSAMDWRQHTSKGKIAMLAAPQGVIRAKTRVTRTPRPAHLWQRP 308

RESULT 13

[illegible]

1 MAVGSKSFATSTLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRIQQ 60
 14 YAVGSKSFATSTLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRIQQ 73
 61 LEMKTRQYAGSQMHEPAPAFQEVAMADHAPAYADHLEGPAMDVRETRYLTLDOTLR 120
 74 LRLTLAAFEAGMOPDPAFAFQEVAMADHAPAYADHLEGPAMDVRETRYLTLDOTLR 133
 121 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180
 134 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 193
 181 EEEGLTKANYAAPENRQALSRAGLVREAPYVSSMAGLAQLPLRSAAIAATKQVYR 240
 194 ODAGLAPENYAARENRPALARWR-RLTDAEPYVYSSQAGLHDLRRRSAAIAATARSVYR 252
 241 KIGVKVEAGKQAMDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 296
 253 EIGIKVKAAGGSANDRRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 309

RESULT 15

WM70124

ABM70124 standard; protein; 309 AA.

ABM70124;

20-NOV-2003 (first entry)

Photorhabdus luminescens protein sequence #3221.

Antibacterial; fungicide; insecticide; polymorphism; genetic analysis; detection; food; gene expression; plant; animal; microorganism; toxin; antibiotic; biopesticide; virulence factor; disease model; plague; whooping cough.

Photorhabdus luminescens.

WC200294867-A2.

28-NOV-2002.

07-FEB-2002; 2002WO-IB003040.

07-FEB-2001; 2001FR-00001659.

(INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.

Duchaud E, Taouit S, Glaser P, Frangeul L, Kunst P, Danchin A, Buchrieser C;

WPI; 2003-148459/14.

Genomic sequence of Photorhabdus luminescens and encoded polypeptides, useful e.g. as therapeutic antimicrobials and agricultural pesticides.

Claim 2; SEQ ID NO 3221; 1205pp; French.

The invention relates to the isolation of genes and their encoded proteins from Photorhabdus luminescens. The isolated sequences are sources of probes and primers for detecting the genome of P. luminescens and related species; to study polymorphisms; for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than P. luminescens and are able to alter response or sensitivity to toxins and antibiotics produced by P. luminescens. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The CC genes, proteins, vectors containing the genes and Ab are also useful CC therapeutically (to treat microbial infection by bacteria or fungi that CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as CC biopesticides. Other uses of the genes and the proteins are as virulence CC factors and for identifying targets of human diseases for which P. CC luminescens is a model (particularly plague and whooping cough). This CC sequence represents one of the isolated P. luminescens proteins

XX Sequence 309 AA;

Query Match 58.9%; Score 904; DB 6; Length 309;

Best Local Similarity 58.6%; Pred. No. 7.6e-95;

Matches 173; Conservative 42; Mismatches 80; Indels 0; Gaps 0;

QY 1 MAVGSKSFATSTLFDKATRSVLMYAWCRHCDVDDQTLGFHADQPSQMEQRIQQ 60

DB 13 MEQSKSFASVTRFLPDTATRHSTWMLYAWCRVCDIIDQELGRQISSVDKYSAREKIQM 72

QY 61 LEMKTRQYAGSQMHEPAPAFQEVAMADHAPAYADHLEGPAMDVRETRYLTLDOTLR 120

DB 73 LQYLTQAYDGLPMTPEPAPAFQTVALSNEIPQQAPEHLEGFANDVLCBPYRTLDOTLR 132

QY 121 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

DB 133 YCYHVAGVYVGLMMAQIMGVDRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 192

QY 181 EEEGLTKANYAAPENRQALSRAGLVREAPYVSSMAGLAQLPLRSAAIAATKQVYR 240

DB 193 HQEGLMPDPTLIYTNRPALARVASRLIVEASVYTSALTGLVGLPLRSAAVIAAHGIYR 252

QY 241 KIGVKVEAGKQAMDRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 295

DB 253 EIGIKVKAAGGSANDRRHROSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 307

Search completed: February 29, 2004, 14:44:06

Job time : 39.0969 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on on: February 29, 2004, 14:33:49 ; Search time 40.4556 Seconds
(without alignments)
3837.172 Million cell updates/sec

title: US-09-941-947A-32
fect score: 2606
quence: 1 MKPTTVIGAGFGLAIALRL.....IPGVIGSAKATAGLMLLEDLI 492

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

itabase : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2606	100.0	492	2	Q8GCSI
2	2453	94.1	492	2	Q47845
3	2298	88.2	492	2	Q8VUJ5
4	1749.5	67.1	506	2	Q9KIX2
5	1713	65.7	504	2	Q93CI7
6	1627	62.4	501	2	Q9RLH4
7	1595	61.2	494	2	P94790
8	1415	54.3	526	2	Q06757
9	1099.5	42.2	511	2	Q9JP98
10	1093.5	42.0	513	2	Q840T3
11	1051	40.3	548	16	Q9RW08
12	984.5	37.8	514	2	Q8KZ28
13	752	28.9	517	2	Q9ZGE4
14	706.5	27.1	511	16	Q8RNM0
15	693.5	26.6	537	16	Q7UF97
16	688.5	26.4	498	16	Q8WS3

17	638.5	24.5	507	2	Q9EXL0	Q9exl0 streptomyce
18	636	24.4	588	2	Q83X01	Q83x01 streptomyce
19	634.5	24.3	534	16	Q7UUB6	Q7ufb6 rhodopirell
20	631.5	24.2	579	3	Q9Y7H8	Q9y7h8 mucor circi
21	629.5	24.2	523	16	Q9RIX9	Q9rix9 streptomyce
22	624	23.9	502	16	Q99R76	Q99r76 staphylococ
23	624	23.9	502	16	Q8NUO6	Q8nuu6 staphylococ
24	617	23.7	506	2	Q07855	Q07855 staphylococ
25	616.5	23.7	513	16	Q93HP2	Q93hp2 streptomyce
26	604	23.2	499	16	Q8YKN8	Q8ykn8 anabaena sp
27	592	22.7	499	2	Q44318	Q44318 anabaena sp
28	568.5	21.8	499	2	Q9K566	Q9k566 mycobacteri
29	567	21.8	490	16	Q8Y8A8	Q8y8a8 listeria mo
30	562	21.6	582	3	Q13506	Q13506 phaffia rho
31	557	21.4	582	3	Q72858	Q72858 phaffia rho
32	549.5	21.1	514	17	Q27835	Q27835 methanobact
33	543.5	20.9	579	3	Q8X0Z0	Q8x0z0 gibberella
34	541.5	20.8	512	17	Q9HPD8	Q9hpd8 halobacteri
35	534	20.5	544	2	Q8KFP34	Q8kfp34 agromyces m
36	533	20.5	494	16	Q8ENM2	Q8enm2 oceanobacil
37	530	19.6	499	16	Q8XK44	Q8xk44 clostridium
38	489	18.8	488	16	Q92D21	Q92d21 listeria in
39	482	18.5	497	16	Q99R73	Q99r73 staphylococ
40	482	18.5	497	16	Q8NUO3	Q8nuu3 staphylococ
41	466	17.9	528	2	Q9KK84	Q9kk84 brevibacter
42	461	17.7	487	2	Q7WT75	Q7wt75 marine bact
43	460	17.7	488	2	Q7WT72	Q7wt72 marine bact
44	433.5	16.6	536	17	Q9HP88	Q9hp88 halobacteri
45	429	16.5	538	16	Q8FRW8	Q8frw8 corynebacte

ALIGNMENTS

RESULT 1
ID Q8GCSI PRELIMINARY; PRT; 492 AA.
AC Q8GCSI;
DT 01-MAR-2003 (TREMREL. 23, Created)
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Phytoene desaturase.
GN CRTI.
OS Pantoea stewartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Pantoea.
NCBI_TaxID=62369;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 8200;
RA deSouza W.L., Kollmann S.R., Schroeder W.A.;
RT "Carotenoid Biosynthesis (WO 02/079395 A2).";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY166713; AAN85599.1;
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac.phytoene dh.
DR Pfam; PF01593; Amino oxidase; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 492 AA; 54836 MW; F3DCD224547A5FBA CRC64;

Query Match 100.0%; Score 2606; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.8e-203; Indels 0;
Matches 492; Conservative 0; Mismatches 0; Gaps 0;

QY 1 MKPTTVIGAGFGLAIALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTFDAGTVITDP 60

Db 1 MKPTTVIGAGFGLAIALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTFDAGTVITDP 60

QY 61 SAIELEFLAGKQLKDYVELLPVTPFYLCHESKVFYNDQALQALNIOCFPRDVG 120


```
b 61 SAIEELFALAGKQKDYVELLPVTFFYRLCWESGKVFYNDNDQAQLEAQIQOQFNPRDVAG 120
y 121 YRAFIDYSAVFNVEGYLKLGTVPFLSFKDMRAAPOLAKLQAWRSVSKVAGYIEDEHLR 180
b 121 YRAFIDYSAVFNVEGYLKLGTVPFLSFKDMRAAPOLAKLQAWRSVSKVAGYIEDEHLR 180
y 181 QAFSFLSLVGGNPFATSSITLIIHALERWGVWPPRGGTGALVNGMIKLFQDLGGEVVL 240
b 181 QAFSFLSLVGGNPFATSSITLIIHALERWGVWPPRGGTGALVNGMIKLFQDLGGEVVL 240
y 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVSNADVVHTYRDLSSQHPAAAKQAKLQSK 300
b 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVSNADVVHTYRDLSSQHPAAAKQAKLQSK 300
y 301 RMSNSFLVLYFGLNHHDDQLAHTVCFGRYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
b 301 RMSNSFLVLYFGLNHHDDQLAHTVCFGRYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
y 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVGPRLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
b 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVGPRLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
y 421 PFDFRDELINAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
b 421 PFDFRDELINAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
y 481 KATAGLMLEDLI 492
b 481 KATAGLMLEDLI 492

RESULT 2
#47845 PRELIMINARY; PRT; 492 AA.
#C Q47845;
#T 01-NOV-1996 (TrEMBLrel. 01, Created)
#Y 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
#E Phytoene dehydrogenase.
#N CRTII.
#S Pantoea agglomerans.
#X Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
#C Enterobacteriaceae; Pantoea.
#X NCBI_TaxID=549;
#X [1]
#P SEQUENCE FROM N.A.
#X MEDLINE=93371414; PubMed=8395826;
#Y Liu S.T.;
#T "Carotenoid-biosynthesis genes as a genetic marker for the purpose of
#T gene cloning.";
#T Biochem. Biophys. Res. Commun. 195:259-263(1993).
#N [2]
#P SEQUENCE FROM N.A.
#X MEDLINE=94236237; PubMed=8180698;
#Y To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
#Y Chang Y.S., Liu S.T.;
#T "Analysis of the gene cluster encoding carotenoid biosynthesis in
#T Erwinia herbicola Eho13.";
#T Microbiology 140:331-339(1994).
#X EMBL; M90688; AAA21263.1; -
#Y FIR; S52586; S52586.
#X GO; GO:0016491; F:oxidoreductase activity; IEA.
#X GO; GO:0009058; P:biosynthesis; IEA.
#X GO; GO:0006118; P:electron transport; IEA.
#X InterPro; IPR002937; Amino oxidase.
#X InterPro; IPR008150; Bac phytoene dh.
#X InterPro; IPR008151; Phytin dehydro.
#X Pfam; PF01593; Amino oxidase; 1.
#X ProDom; PD139017; Phytin dehydro; 1.
#X PROSITE; PS00982; PHYTOENE DH; 1.
#Q SEQUENCE 492 AA; 55010 MW; 2D65E1A2A32D0635 CRC64;

Query Match 94.1%; Score 2453; DB 2; Length 492;
```

```
Best Local Similarity 93.7%; Pred. No. 7.9e-191;
Matches 461; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

QY 1 MKPTTIVGAGFGGLALAIRLQAAGIPVLLLEQRDKPGRAYVYQEGFTDAGTVIDP 60
DB 1 MKPTTIVGAGFGGLALAIRLQAAGIPVLLLEQRDKPGRAYVYQEGFTDAGTVIDP 60
QY 61 SAIEELFALAGKQKDYVELLPVTFFYRLCWESGKVFYNDNDQAQLEAQIQOQFNPRDVAG 120
DB 61 SAIEELFALAGKQKDYVELLPVTFFYRLCWESGKVFYNDNDQAQLEAQIQOQFNPRDVAG 120
QY 121 YRAFIDYSAVFNVEGYLKLGTVPFLSFKDMRAAPOLAKLQAWRSVSKVAGYIEDEHLR 180
DB 121 YRAFIDYSAVFNVEGYLKLGTVPFLSFKDMRAAPOLAKLQAWRSVSKVAGYIEDEHLR 180
QY 181 QAFSFLSLVGGNPFATSSITLIIHALERWGVWPPRGGTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSFLSLVGGNPFATSSITLIIHALERWGVWPPRGGTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVSNADVVHTYRDLSSQHPAAAKQAKLQSK 300
DB 241 NARVSHMETVGDKIQAQVLEDGRRFETCAVSNADVVHTYRDLSSQHPAAAKQAKLQSK 300
QY 301 RMSNSFLVLYFGLNHHDDQLAHTVCFGRYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
DB 301 RMSNSFLVLYFGLNHHDDQLAHTVCFGRYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVGPRLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVGPRLRDRIFDYLEOHYMPGLRSQLVTHRMFT 420
QY 421 PFDFRDELINAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDELINAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
QY 481 KATAGLMLEDLI 492
DB 481 KATAGLMLEDLI 492

RESULT 3
Q8VUJ5 PRELIMINARY; PRT; 492 AA.
AC Q8VUJ5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CrtI protein.
GN CrtI.
OS Pantoea agglomerans pv. milletiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=182454;
OX [1]
RN SEQUENCE FROM N.A.
RA Kamlunten H., Hirata R.;
RT "Isolation and characterization of carotenoid biosynthesis genes from
RT Pantoea agglomerans pv. milletiae Wist 801.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076662; BAB79603.1; -
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phytin dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 492 AA; 54802 MW; C3EE7DB0D96E3095 CRC64;

Query Match 88.2%; Score 2298; DB 2; Length 492;
Best Local Similarity 88.0%; Pred. No. 3.2e-178;
```

Matches 432; Conservative 26; Mismatches 33; Indels 0; Gaps 0;

1 MKPTTVIGAGGGGLAALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTEDAGPTVITDP 60
 1 MNRVTIVAGGGGLAALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTEDAGPTVITDP 60
 61 SAIELFALAGKQKDYVELLPVTPFFRLCWESKVFYNDQALQAEQIQOQFNPRDVAG 120
 61 SAIELFALAGKQKDYVELLPVTPFFRLCWESKVFYNDQALQAEQIQOQFNPRDVAG 120
 61 SAIELFTLACKKLSDYVELMPVFPFRLCWESKVFYNDQALQAEQIQOQFNPRDVAG 120
 121 YRAFIDYSRAVFNQGYLKLGTVPFLSKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 180
 121 YRFLAYSRAVFAEGYLKLGTVPFLSKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 180
 181 QAFPSHLLVGNPFATSSIIYTLHALEREWGVPFRGGTGALVNGMKLFDLGGSEVL 240
 181 QAFPSHLLVGNPFATSSIIYTLHALEREWGVPFRGGTGALVNGMKLFDLGGSEVL 240
 241 NARYSHMETVGDKIQAQVLEDGRPFETCAVASNADVVHTYRDLLSQHPAAKQAKLQSK 300
 241 NASVARLETQENRITAVHLKDGVRFPTRAVASNADVVHTYRELLSQHPASQAQGRSLQNK 300
 301 RMSNSLFVLYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTDP 360
 301 RMSNSLFVLYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTDP 360
 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMFGLSQLVTHRMFT 420
 361 SLAPGCGSYVYLAAPVPHLGTADLDWAVEGPRLDRIFDYLEQHYMFGLSQLVTHRIFT 420
 421 PFDREDELNAWQGSFAFSEPTLTOSANFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 421 PFDREDELNAWQGSFAFSEPTLTOSANFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 481 KATAGLMLEDL 491
 481 KATAGLMLEDL 491

RESULT 4
 KIX2 PRELIMINARY; PRT; 506 AA.
 Q9KIX2;
 01-OCT-2000 (TREMELrel. 15, Created)
 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 Phytoene desaturase.
 CRTI.
 Brachyribzobium sp. ORS278.
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.
 NCBI_TaxID=114615;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-ORS278;
 MEDLINE=20309720; PubMed=10851005;
 Hannibal L., Lorquin J., Angles d'Ortoli N., Garcia N.,
 Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;
 "Isolation and characterization of the canthaxanthin biosynthesis
 genes from the photosynthetic bacterium Brachyribzobium sp. strain
 ORS278";
 J. Bacteriol. 182:3850-3853 (2000).
 EMBL; AF218415; AAF78201.1; -;
 GO; GO:0016491; P:oxidoreductase activity; IEA.
 GO; GO:0009058; P:biosynthesis; IEA.
 GO; GO:0006118; P:electron transport; IEA.
 InterPro; IPR002937; Amino oxidase.
 InterPro; IPR008150; Bac_phytoene dh.
 InterPro; IPR008151; Phytin dehydro.
 Pfam; PF01593; Amino oxidase; 1.
 ProDom; PD139017; Phytin dehydro; 1.
 ProDom; PD139017; Phytin dehydro; 1.
 PROSITE; PS00982; PHYTOENE DH; 1.
 SEQUENCE 506 AA; 56223 MW; 435F949EGFDEED3 CRC64;

Query Match 67.1%; Score 1749.5; DB 2; Length 506;
 Best Local Similarity 67.8%; Pred. No. 1.4e-133;
 Matches 332; Conservative 55; Mismatches 102; Indels 1; Gaps 1;

QY 2 KPTTVIGAGGGGLAALRLQAAGIPVLLLEQRDKPGGRAYVYQEGFTEDAGPTVITDP 61
 DB 9 KPAVVGSGGGLSLARLKAGIATTLVEKRDIFGGRAYVYQEGFTEDAGPTVITDP 68
 QY 62 AIELFALAGKQKDYVELLPVTPFFRLCWESKVFYNDQALQAEQIQOQFNPRDVAG 121
 DB 69 CLQELFALSGKLENYVELMPVSPFYQLRWEDGATFDVNDQALQAEQIAAFCPADVGY 128
 QY 122 RAFPIDYSRAVFNQGYLKLGTVPFLSKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 181
 DB 129 RFRFSYSERLEEGYVGLGHVFPDFFRSMVAVAPOLVALQSYRSVYSKVAGYIEDEHLR 188
 QY 182 AFSHLSLLVGNPFATSSIIYTLHALEREWGVPFRGGTGALVNGMKLFDLGGSEVL 240
 DB 189 AFSHLSLLVGNPFATSSIIYTLHALEREWGVPFRGGTGALVNGMKLFDLGGSEVL 248
 QY 241 NARYSHMETVGDKIQAQVLEDGRPFETCAVASNADVVHTYRDLLSQHPAAKQAKLQSK 300
 DB 249 STSVSRIETANGRVSAAVVAEDGRFADIVASNADVVHTYRDLLKDEPLARPTAQLARK 308
 QY 301 RMSNSLFVLYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTDP 360
 DB 309 RFSNSLFVLYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTDP 368
 QY 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMFGLSQLVTHRMFT 420
 DB 369 SLAPGCGSYVYLAAPVPHLGTADLDWAVEGPRLDRIFDYLEQHYMFGLSQLVTHRIFT 428
 QY 421 PFDREDELNAWQGSFAFSEPTLTOSANFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 DB 429 PFDREDELNAWQGSFAFSEPTLTOSANFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 488
 QY 481 KATAGLMLEDL 490
 DB 489 KATAGLMLEDL 498

RESULT 5
 Q93CI7 PRELIMINARY; PRT; 504 AA.
 AC Q93CI7;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Phytoene dehydrogenase.
 GN CRTI.
 OS Xanthobacter sp. (strain Py2).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Hyphomicrobiaceae; Xanthobacter.
 OX NCBI_TaxID=78245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Py2;
 RA Larsen R.A., Metcalf W.W.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF408848; AAL02000.1; -;
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR002937; Amino oxidase.
 DR InterPro; IPR008150; Bac_phytoene dh.
 DR InterPro; IPR008151; Phytin dehydro.
 DR Pfam; PF01593; Amino oxidase; 1.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR ProDom; PD139017; Phytin dehydro; 1.
 PROSITE; PS00982; PHYTOENE DH; 1.
 SQ SEQUENCE 504 AA; 55556 MW; CA848D11A607A828 CRC64;

Query Match 65.7%; Score 1713; DB 2; Length 504;

```

Best Local Similarity 66.0%, Pred. No. 1.3e-130;
Matches 324; Conservative 56; Mismatches 111; Indels 0; Gaps 0

1y 1 MKPTTVCAGFGGLALAIRLQAAGIPVLLLEQRDKPGRAYVYQEGFTFDAGPTVTD 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 9 IRRAAVIGSGFGLSLAIRLQAAGIRTTVEQRDKPGRAYVYEQDFTFDGPTVTD 68

1y 61 SAISELPALAGKOLKQVVELLPVTPTFVRLCWESKGVNTDNOAQLEAQIQPNPRDVAG 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 69 TCLEVEFAAAGKLSDYVELMPVSPFYRLWPDGRQFVYVNOQAALDAQIATFDKADVBG 128

1y 121 YRAFLDYSRAVFNEGYKLGTVPFLSPKDMRAAPOLAKLQAWRSVYSKVAGYVDEHLR 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 129 YRRFLASQAVFEGYKLGAVPFLPSSMKXAPKLVLEAWRSYVAMVSPFIRDDHLR 188

1y 181 QAFSFHLLVGGNPFATSSYITLIHALERWGVMPFRGGTGALVNGWIKLFDLGGVVVL 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 189 QAFSFHLLVGGNPFSTSSIVALIHALERKGVFPFRGGTGALVRGWKLFITDLGGEIRL 248

1y 241 NARYSHMETVGDKIQAVOLEDRRFETCAVASNADVHTVTRDLLSHQPAAAKQAQKQSK 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 249 TSPDEIVWEGQRATAVLKSGETLPFDLVASNADVHTVTRHLLRGAARGSEGGARLAKT 308

1y 301 RMSNSLFVYGLNHHHDLQAHTTFCGPRYRBLIHEIFNHDGLAEDFSYLHAPCVTDP 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 309 RHSNSLFTVYFGAERTWDHLQHTVFLGPRYRGLVDEIPKGENLPDFFSLYLHAPVTDK 368

1y 361 SLAPEGCSYTLAPVPHLGTANLDRAVEGPRLDRIFDYLEQHYMPGLRSQIVTRHMT 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 369 SLAPEGCTAFYULSPVPHLGKADIDMEVEGFLYRDLIAHLBERLLPGLRDLVTSRILT 428

1y 421 PFDPRDELNAWQGSAPVEPLITQSAWFRPHNRKIDNLVYLGAGTHPGAGIPGVIGSA 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1b 429 PFGPRDELSAQGSAPVEPLITQSAWFRPHNRDAKIANLYFAGAGTHPGAGVPGWVISA 488

1y 481 KATAGLMLDL 491
1b 489 KATAGLILADI 499

```

Best Local Similarity	63.5%	Pred. No. 1-2e-123;	
Matches 311;	Conservative 58;	Mismatches 121;	Indels 0; Gaps 0;
QY	2	KPTTVIGAGFGLALAIRLQAAGIPVLLLEBQRDKPGGRAYVYQGQFTFDAGPTVITDPS	61
Db	9	KTAIVIGAGFGLALAIRLQAGIATTLVBARDKPGGRAYVWHQDGHVFDAGPTVITDPS	68
QY	62	ALIELFALAGKQADXYVELLPVTPFVFLCWESGKVFNYNDQALQAEQAQQFNPRDVAGY	121
Db	69	AKLELWALTQGDMDVTLMPVSPFTRLMWPGSKVFDYVNEADQLERQIAQNPDDLEGY	128
QY	122	RAFLDYSRAVFNEGYLKGLTVPFLSPFKMLRAAPQLAKLQAMRSVYKVGAYIEDEHLRQ	181
Db	129	RRFRDYAEVYQGVKLGTVPFLLKQMLKAAPALMKLEAYNSVHAKVATFIKDPYLRQ	188
QY	182	AFSPHSLVGCNPATSSIVTLHALEREWGVPFPGGTGALVNGMIKLFODLGGEVVLN	241
Db	189	AFSVHTLVGCNPFSTSSIVLHALERRGVWFARGGTNQLVAGHMVALFERPGQOMLIN	248
QY	242	ARVSHMETVGDGKIQAVALQEDGRFPETCAVASNADVVHTVTRDLLSQHPAAAKQKLOSKR	301
Db	249	AKVARIETEGARTTGVTLADGRSLRADMVASNGDVHNYRDLGLGHTARGSKRAKSLDRK	308
QY	302	MENSLFVYVFGNLHHHDLQAHVTVCGSPRYRELIIHFNHHDGLAEDFSLYLHAPCVITDPS	361
Db	309	WMSLFLVHLFGLRBPXDIARHTILFGPRYKELVNEIFKGPKLAEDFSLYXHSPTTDDP	368
QY	362	LAPEGCGYVVLAVPVLHGTANLDWAVGEPRLDRIFDYLEOHVMPGLRSQLVTHRMETP	421
Db	369	WAPPQWSTHYVLAVPPLHSRAEIDWAVEGPRYADRILAFLEERLIPLNLANLTRTITFT	428
QY	422	EDFRDELNAWQGSFAFSVEPILTSQAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAK	481
Db	429	ADFASELNAHGSFAFSVEPILTSQAMFRPHNRDKTIRNPFYLVGAGTHPGAGIPGVIGSAK	488
QY	482	ATAGLWLEDL	491
Db	489	ATAQVWLSDL	498
RESULT 7			
ID	P94790	PRELIMINARY;	PRT; 494 AA.
AC	P94790;		
DT	01-MAY-1997 (TRENBLrel. 03, Created)		
DT	01-MAY-1997 (TRENBLrel. 03, Last sequence update)		
DT	01-JUN-2003 (TRENBLrel. 24, Last annotation update)		
DE	Phytoene desaturase.		
DE	Phytoene		
CS	CRTI.		
GN	Flavobacterium sp. ATCC 21588.		
OC	Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;		
OC	Flavobacteriaceae; Flavobacterium.		
NC	NCBI_TaxID=50286;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=R1534.		
RC	MEDLINE=97186694; PubMed=9034310;		
RA	Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,		
RA	van Ioon A.P.;		
RT	"Isolation and characterization of the carotenoid biosynthesis genes		
RT	of flavobacterium sp. strain R1534.";		
RL	Gene 185:35-41[1997].		
DR	EMBL, U62808; AAC44850.1; -		
DR	GO; GO:0016451; F:oxidoreductase activity; IEA.		
DR	GO; GO:0009058; P:biosynthesis; IEA.		
DR	GO; GO:006118; P:electron transport; IEA.		
DR	InterPro; IPR002937; Amino oxidase.		
DR	InterPro; IPR008150; Bac.phytoene dh.		
DR	InterPro; IPR008151; Phytn.dehydro.		
DR	Pfam; PF01593; Amino oxidase; 1.		
DR	ProDom; PD139017; Phytn.dehydro; 1.		
DR	PROSITE; PS00982; PHYTOENE DH; 1.		
SQ	SEQUENCE 494 AA; 5441 MW; CBF03BB39C1FC4D0 CRC64;		

```

RESULT 6
19RLH4
1D Q9RLH4 PRELIMINARY; PRT; 501 AA.
1C Q9RLH4;
1T 01-MAY-2000 (TrEMBLrel. 13, Created)
1T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
1T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
1E Phycoene desaturase.
1N CRTI.
1S Paracoccus marcusii.
1C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
1C Rhodobacteraceae; Paracoccus.
1X NCBI TaxID=59779;
1N [1]_taxID=59779;
1P SEQUENCE FROM N.A.
1C STRAIN=WHI;
1A Harker M., Hirschberg J.;
1T "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
1T WHI.";
1L Submitted (SRP-1997) to the EMBL/GenBank/DBJ databases.
1R EMBL; Y15112; CAB56062.1; ...
1R GO; GO:0016491; P:oxidoreductase activity; IEA.
1R GO; GO:0009058; P:biosynthesis; IEA.
1R GO; GO:0006118; P:electron transport; IEA.
1R InterPro; IPR002937; Amino oxidase.
1R InterPro; IPR008150; Bac_phytoene dh.
1R InterPro; IPR008151; Phyt_n dehydro.
1R Pfam; PF01593; Amino oxidase; 1.
1R ProDom; PD133017; Phyt_n dehydro; 1.
1R PROSITE; PS00982; PHYTOENE DH; 1.
1Q SEQUENCE 501 AA; 55109 MW; 70FAAED92A0E8D16 CRC64;

Query Match 62.4%; Score 1627; DB 2; Length 501;

```

```

Query Match      61.2%; Score 1595; DB 2; Length 494;
Best Local Similarity 61.9%; Pred. No. 4, 9e-121;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

Y 1 MKFTTVIGAGGGLAALRIQAGIPVLLLEORDKPGRAYVYQGGTFPDAGPTVITDPA 60
b 1 MSSAIVIGAGGGLAALRIQAGIATTVEARDKPGGRAYVYQGGTFPDAGPTVITDPA 60
Y 61 SAIEELFALAGKQKDYVELLPVTPPYRLCWESGVFNVDNDQAGLEAQIQOENPRDVAG 120
b 61 DSLRELWALSGQPMERDVTLLPVSPYRLTWADGSEFYVNDDELRIQVASEFNPADVDG 120
Y 121 YRAPLDYSRAVNEGYKLGTVPFVSKDMLRAAPQALAKQAWRSVYKVGAYTDEHLR 180
b 121 YRPFHDYAEVYREGYKLGTVPFVSKDMLRAAPQALAKQAWRSVYKVGAYTDEHLR 180
Y 181 QAFSPHLLVGNPPATSIYTLIHALERWGVWPRGGTGALVNGMKILFQDLGGSEVVL 240
b 181 QAFSPHLLVGNPPATSIYTLIHALERWGVWPRGGTGALVNGMKILFQDLGGSEVVL 240
Y 241 NARVSHMETVGKIQAVLEDDRRFETCAVASNADVHTYRDLLSQHPAAAKQAKLQSK 300
b 241 NARVTRIDTEGORATGVTLDDGRLQADTVASNGDVMSYRDLLGHTRRGRTKAILNRQ 300
Y 301 RMSNSLFLVYFGLNHHDDLAHTVCFGRPYRELIIHEIFNHDGLAEDFSLYLHAPCVTD 360
b 301 RMSNSLFLVYFGLNHHDDLAHTVCFGRPYRELIIHEIFNHDGLAEDFSLYLHAPCVTD 360
Y 361 SLAPEGCGSYVYVLAPVPHLGTANLDMWAVEGPRLRDRIFDYLEQHTMGLRSLQVTHRMFT 420
b 361 SLAPEGCMSTHYVLAPVPHLGTANLDMWAVEGPRLRDRIFDYLEQHTMGLRSLQVTHRMFT 420
Y 421 PFDPRDELNAWGSAPSEPIITQSAWFRPHNRDKHIDNLVYGAGTHPGAGIPGVIGTGA 480
b 421 PADFSTLSAHGSAFSEPIITQSAWFRPHNRDKHIDNLVYGAGTHPGAGIPGVIGTGA 480
Y 481 KATAGLMLDL 491
b 481 KATAQWMLSDL 491

RESULT 8
Q6757 PRELIMINARY; PRT; 526 AA.
01-JUL-1997 (TrEMBLrel. 04, Created)
01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
Phytoene desaturase.
CITI.
Erythrocyte longus.
Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
Sphingomonadaceae; Erythrocyte.
NCBI_TaxID=1044;
[1]
SEQUENCE FROM N.A.
MEDLINE=97311406; PubMed=9168123;
Strain=ochi101;
Matsumura H., Takeyama H., Hsuakabe E., Burgess J.G., Matsunaga T.;
"Cloning, sequencing and expression of the carotenoid biosynthesis gene,
lycopene cyclase and phytoene desaturase, from the aerobic
phototrophic bacterium Erythrocyte longus sp. strain Och 101 in
Escherichia coli."
Gene 189:169-174(1997).
EMBL; D83514; BAA20276.1;
GO: 0016491; P:oxidoreductase activity; IEA.
GO: 0009058; P:biosynthesis; IEA.
GO: 0006118; P:electron transport; IEA.
InterPro: IPR002937; Amino_oxidase.
InterPro: IPR008150; Bac phytoene dh.
InterPro: IPR008151; Phytn dehydro.
Pfam: PF01593; Amino_oxidase; 1.

DR ProDom: PD139017; Phytn dehydro; 1.
DR PROSITE: PS00982; PHYTOENE DH; 1.
SQ SEQUENCE 526 AA; 58469 Kw; 66866FP366P652232 CRC64;

Query Match      54.3%; Score 1415; DB 2; Length 526;
Best Local Similarity 55.1%; Pred. No. 2, 4e-106;
Matches 268; Conservative 77; Mismatches 141; Indels 0; Gaps 0;

QY 6 VIGAGGGGLAALRIQAGIPVLLLEORDKPGRAYVYQGGTFPDAGPTVITDPA 65
b 6 VIGAGGGGLAALRIQAGIATTVEARDKPGGRAYVYQGGTFPDAGPTVITDPA 65
QY 38 VIGAGGGGLAALRIQAGIATTVEARDKPGGRAYVYQGGTFPDAGPTVITDPA 97
b 38 VIGAGGGGLAALRIQAGIATTVEARDKPGGRAYVYQGGTFPDAGPTVITDPA 97
QY 66 LFLAGKQKDYVELLPVTPPYRLCWESGVFNVDNDQAGLEAQIQOENPRDVAGYRAFL 125
b 66 LFLAGKQKDYVELLPVTPPYRLCWESGVFNVDNDQAGLEAQIQOENPRDVAGYRAFL 125
QY 98 LWBTLGHTDISDVELMKVHPFVRLNPDGTFDYSNVDEELNAETAKLNPDDVIGYKEL 157
b 98 LWBTLGHTDISDVELMKVHPFVRLNPDGTFDYSNVDEELNAETAKLNPDDVIGYKEL 157
QY 126 DYSAVFNEGYKLGTVPFVSKDMLRAAPQALAKQAWRSVYKVGAYTDEHLQAFSP 185
b 126 DYSAVFNEGYKLGTVPFVSKDMLRAAPQALAKQAWRSVYKVGAYTDEHLQAFSP 185
QY 158 EYSARVHEBTVKLGTVPFVSKDMLRAAPQALAKQAWRSVYKVGAYTDEHLQAFSP 217
b 158 EYSARVHEBTVKLGTVPFVSKDMLRAAPQALAKQAWRSVYKVGAYTDEHLQAFSP 217
QY 186 HSLVGNPPATSIYTLIHALERWGVWPRGGTGALVNGMKILFQDLGGSEVVLNARYS 245
b 186 HSLVGNPPATSIYTLIHALERWGVWPRGGTGALVNGMKILFQDLGGSEVVLNARYS 245
QY 218 HTLVGSGPMKTSIAIYALIKLEKDGVMWARGGTNRLIAGVVRHFERLGGMTRIGDPV 277
b 218 HTLVGSGPMKTSIAIYALIKLEKDGVMWARGGTNRLIAGVVRHFERLGGMTRIGDPV 277
QY 246 HMETVGDKIQAVLEDDRRFETCAVASNADVHTYRDLLSQHPAAAKQAKLQSKMSNS 305
b 246 HMETVGDKIQAVLEDDRRFETCAVASNADVHTYRDLLSQHPAAAKQAKLQSKMSNS 305
QY 278 QVHTQGTQKATEVETSKGWERFDVAVCSNADIMHYSKELLGESDRGKRYAKSLKSYSP 337
b 278 QVHTQGTQKATEVETSKGWERFDVAVCSNADIMHYSKELLGESDRGKRYAKSLKSYSP 337
QY 306 LFVLYFGLNHHDDLAHTVCFGRPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
b 306 LFVLYFGLNHHDDLAHTVCFGRPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
QY 338 LFVYFGLSGMPGIAHMLFGPRYKELVDDIYKHGVLQDFSIYLLHHTVTDPSMAPK 397
b 338 LFVYFGLSGMPGIAHMLFGPRYKELVDDIYKHGVLQDFSIYLLHHTVTDPSMAPK 397
QY 366 GCGSYTVLAPVPHLGTANLDMWAVEGPRLRDRIFDYLEQHTMGLRSLQVTHRMFTFFDR 425
b 366 GCGSYTVLAPVPHLGTANLDMWAVEGPRLRDRIFDYLEQHTMGLRSLQVTHRMFTFFDR 425
QY 398 GMSFTYALVFAHLGKMPIDMDVDEGPKFEKALIDETGRRLIPDIHDIIVTKESYAPKDF 457
b 398 GMSFTYALVFAHLGKMPIDMDVDEGPKFEKALIDETGRRLIPDIHDIIVTKESYAPKDF 457
QY 426 DELNAWGSAPSEPIITQSAWFRPHNRDKHIDNLVYGAGTHPGAGIPGVIGTGA 485
b 426 DELNAWGSAPSEPIITQSAWFRPHNRDKHIDNLVYGAGTHPGAGIPGVIGTGA 485
QY 458 ADLNAWGSAPSEPIITQSAWFRPHNRDKHIDNLVYGAGTHPGAGIPGVIGTGA 517
b 458 ADLNAWGSAPSEPIITQSAWFRPHNRDKHIDNLVYGAGTHPGAGIPGVIGTGA 517
QY 486 LMLEDL 491
b 486 LMLEDL 491
QY 518 LMLEDL 523
b 518 LMLEDL 523

RESULT 9
Q9JP98 PRELIMINARY; PRT; 511 AA.
AC Q9JP98;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN CITI.
OS Rhodococcus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LI144;
RA Nagashima K.V., Shimada K., Matsuura K.;
RT "Phylogenetic analysis of photosynthetic genes of Rhodococcus
RT gelatinosus: Possibility of horizontal gene transfer in purple
RT bacteria."
RL Photosyn. Res. 36:185-191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LI144;
RX MEDLINE=94132007; PubMed=8300574;
RA Nagashima K.V., Matsuura K., Ohshima S., Shimada K.;
RT "Primary structure and transcription of genes encoding B870 and
RT photosynthetic reaction center apoproteins from Rubrivivax

```

gelatinous.";
 J. Biol. Chem. 269:2477-2484(1994).
 [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=11144;
 RA Igarashi N., Shimada K., Matsuura K., Nagashima K.V.;
 RA "Photosynthetic gene cluster in purple bacterium, Rubrivivax
 XT gelatinous"; (eds.);
 IL (In) Garab G. (eds.);
 IL Photosynthesis;
 IL mechanisms and effects (Proceedings of the 11th international congress
 IL on photosynthesis), pp.4:2889-2892, Kluwer Academic Publishers,
 IL Dordrecht (1999).
 IN [4]
 IC SEQUENCE FROM N.A.
 IC STRAIN=11144;
 IC MEDLINE=20031519; PubMed=10563807;
 JA Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,
 JA Parot P., Vermeig A.;
 JT "Dark aerobic growth conditions induce the synthesis of a high
 JT midpoint potential cytochrome c8 in the photosynthetic bacterium
 T Rubrivivax gelatinous";
 T Biochemistry 38:15238-15244(1999).
 R EMBL; AB034704; BAA94063.1; -;
 R FJ; T5091G; T5091G.
 R GO; GO:0016491; F:oxidoreductase activity; IEA.
 R GO; GO:0009058; P:electron transport; IEA.
 R GO; GO:0006118; P:electron transport; IEA.
 R InterPro; IPR002937; Amino oxidase.
 R InterPro; IPR008150; Bac.phytoene_dh.
 R InterPro; IPR000205; NAD_BS.
 R InterPro; IPR008151; Phytin dehydro.
 R Pfam; PF01593; Amino oxidase; 1.
 R ProDom; PD139017; Phytin dehydro; 1.
 R PROSITE; PS00982; PHYTOENE DH; 1.
 Q SEQUENCE 511 AA; 56818 MW; 95648971960333903 CRC64;
 Query Match 42.2%; Score 1099.5; DB 2; Length 511;
 Best Local Similarity 42.3%; Pred. No. 1e-80;
 Matches 204; Conservative 100; Mismatches 177; Indels 1; Gaps 1;
 Y 6 VIGAGFGGLALAIRLOAGIPVLLSQRDKPGGRAYVYQEGFTFDAGTWTDSALIEE 65
 b 23 VVSGFGGMAAVALAAKGRVTVLEKLDAPGRAYVHRREGHVFDAGTIVTVPLFDE 82
 y 66 LFALAGKQLKDYVELLPVTPFYRLCHWESKVFNYNDQALQAEQIQENPRDVAGYRAFL 125
 b 83 LWLAGRKFSDDIELKSLDFYRIRFDGDDHFDYSGDPARMAEVRRISSPSDAGFERFM 142
 y 126 DYSAVFNEGYKLGTVFPFLSPFKOMLRAPOLAKQAWRSVYKVGATYEDHLRQAFSP 185
 b 143 READQCVELGFTGLGDKAFDTVGDLKAAPLIIRLGRWRSLSHOMVSSHLKHPKLAIA 202
 y 186 HSLVGNPNPATSIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGVVLNARVS 245
 b 203 QSLILGNPFVSMYALYNALERQNGVHMGVGGTGLIRGLVDVFEKGGMTRLKAEVK 262
 y 246 HMETVGDKIQAQVLEGRFETCAVASNADVHTYRDLDSQHPAAAKQAKKLOSKRMNS 305
 b 263 RIEVDNGVATGVTADGERIPADIVVCNGDTGLYKNLVDAERWRKHWTDAIERGHYSG 322
 y 306 LFVLYPGLNHHDDQLAHTVCFGRPRRELIHIEFNHDLGAEFSLYHAPCVTDPSLAPE 365
 b 323 LFVWYFGTDRRYEDVPHMMVLGPRYRELLDDIFRKKLASDFSILHRPTATDSMAPA 382
 y 366 GCSYVTLAPVPHLGTANLDMAVEGRRLDRIDFYLEQHYMPGLRSQVLTWRMFPFDFR 425
 b 383 GCDTFALMPVPHLGSCT-DWTTQAEPRYQSVQEALERTVLPGQLHRLVFCFTPLDFQ 441
 y 426 DELNAGCSAFSVEPILTOSAFWRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAKATAG 485
 b 442 HRLSYKAGFGLEGLPULLQSAVFRPHNRSEVDKLNLFMVGASTHPGAGVGVINSKALLES 501

QY 496 IX 487
 DB 502 VL 503
 RESULT 10
 Q84073 PRELIMINARY; PRT; 513 AA.
 ID Q84073;
 AC Q84073;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Phytoene dehydrogenase.
 GN CRH.
 OS Rhodocyclus gelatinosus (Rhodospseudomonas gelatinosa).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Comamonadaceae; Rubrivivax.
 CX NCBI_TaxID=28068;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SI;
 RA Ouchane S., Steunou A.-S., Astier C.;
 RA "Photosynthesis Gene Regulation in Rubrivivax gelatinosus:
 RT Transcription Factor PpsR is Involved in both Negative and Positive
 RT Control";
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY234385; AAC93135.1; -;
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0009058; P:electron transport; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR002937; Amino oxidase.
 DR InterPro; IPR008150; Bac.phytoene_dh.
 DR InterPro; IPR000205; NAD_BS.
 DR InterPro; IPR008151; Phytin dehydro.
 DR Pfam; PF01593; Amino oxidase; 1.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR PROSITE; PS00982; PHYTOENE DH; 1.
 SQ SEQUENCE 513 AA; 57074 MW; 9C46447BB02F682D CRC64;
 Query Match 42.0%; Score 1093.5; DB 2; Length 513;
 Best Local Similarity 42.1%; Pred. No. 3.2e-80;
 Matches 203; Conservative 101; Mismatches 177; Indels 1; Gaps 1;
 QY 6 VIGAGFGGLALAIRLOAGIPVLLSQRDKPGGRAYVYQEGFTFDAGTWTDSALIEE 65
 DB 23 VVSGFGGMAAVALAAKGRVTVLEKLDAPGRAYVHRREGHVFDAGTIVTVPLFDE 84
 QY 66 LFALAGKQLKDYVELLPVTPFYRLCHWESKVFNYNDQALQAEQIQENPRDVAGYRAFL 125
 DB 85 LWLAGRKFSDDIELKSLDFYRIRFDGDDHFDYSGDPARMAEVRRISSPSDAGFERFM 144
 QY 126 DYSAVFNEGYKLGTVFPFLSPFKOMLRAPOLAKQAWRSVYKVGATYEDHLRQAFSP 185
 DB 145 READQCVELGFTGLGDKAFDTVGDLKAAPLIIRLGRWRSLSHOMVSSHLKHPKLAIA 204
 QY 186 HSLVGNPNPATSIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGVVLNARVS 245
 DB 205 QSLILGNPFVSMYALYNALERQNGVHMGVGGTGLIRGLVDVFEKGGMTRLKAEVK 264
 QY 246 HMETVGDKIQAQVLEGRFETCAVASNADVHTYRDLDSQHPAAAKQAKKLOSKRMNS 305
 DB 265 RIEVDNGVATGVTADGERIPADIVVCNGDTGLYKNLVDAERWRKHWTDAIERGHYSG 324
 QY 306 LFVLYPGLNHHDDQLAHTVCFGRPRRELIHIEFNHDLGAEFSLYHAPCVTDPSLAPE 365
 DB 325 LFVWYFGTDRRYEDVPHMMVLGPRYRELLDDIFRKKLASDFSILHRPTATDSMAPA 384
 QY 366 GCSYVTLAPVPHLGTANLDMAVEGRRLDRIDFYLEQHYMPGLRSQVLTWRMFPFDFR 425
 DB 385 GCDTFALMPVPHLGSCT-DWTTQAEPRYQSVQEALERTVLPGQLHRLVFCFTPLDFQ 443
 QY 426 DELNAGCSAFSVEPILTOSAFWRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAKATAG 485

444 HRLSVKAGAGLEPLLLQSAYPRHRSEDVKNLFWMGASTHPCAGVPGVMSAKALES 503
 486 LM 487
 504 VL 505

RESULT 11
 9RM08 PRELIMINARY; PRT; 548 AA.
 01-XAY-2000 (TREMELrel. 13, Created)
 01-XAY-2000 (TREMELrel. 13, Last sequence update)
 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 1 Phytoene dehydrogenase.
 2 DR0861.
 3 Deinococcus radiodurans.
 4 Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 5 Deinococcaceae; Deinococcus.
 6 NCBI_TaxID=1299;
 7 (1)
 8 SEQUENCE FROM N.A.
 9 STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
 10 MEDLINE=20036896; PubMed=10567266;
 11 White O., Eisen J.A., Heidelberg J.F., Hickey E.X., Peterson J.D.,
 12 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 13 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 14 Vamathevan J.S., Lam P., McDonald L., Antonback T., Zalewski C.,
 15 Makarova K.S., Aravind L., Daly M.J., Minter K.W., Fleischmann R.D.,
 16 Ketchum K.A., Nelson K.B., Salzberg S., Smith H.O., Venter J.C.,
 17 Fraser C.M.;
 18 "Genome sequence of the radioresistant bacterium Deinococcus
 19 radiodurans R1.";
 20 Science 285:1571-1577(1999).
 21 EMBL; AE001940; AF0439.1; -.
 22 PIR; C75486; C75466.
 23 TIGR; DR0861; -.
 24 GO; GO:006118; P:electron transport; IEA.
 25 InterPro; IPR002937; Amino oxidase.
 26 InterPro; IPR008151; Phytin dehydro.
 27 Pfam; PF01593; Amino oxidase; 1.
 28 ProDom; PD139017; Phytin dehydro; 1.
 29 Complete proteome.
 30 SEQUENCE 548 AA; 60188 MW; 80A20729284A9A6C CRC64;

Query Match 40.3%; Score 1051; DB 16; Length 548;
 Best Local Similarity 43.5%; Pred. No. 1e-76;
 Matches 229; Conservative 72; Mismatches 185; Indels 40; Gaps 9;

2 KPTTVIGAGFGLALAIRLQAGIPVLLLEQDKPGGRAYVYQ-EGQFTFDAGPTVITDP 60
 17 KTAIVIGAGFGLALGIRLSIGFDITLRLDGGGRAYQKRTPDGVVDMGPTVLTVP 76
 61 SAIELEFA-----LAGKQKQ-----YVELLPVTPYRLCWES 93
 77 HFIEELFALEDRAGLDADYPPEVLGSEYKEGVGGPHTSRYTLVPLPTFYRVFHD 136
 94 GKVFYNDQQAQLBAIQOQFNPRDVAGRAPLDYRAVFNQGYLKLGTVPFLSKMQLRA 153
 137 GTFYDGDGPSTRRQFAELAPGLAGYERFADABAIERRGFLEGLGYTHFGDVPTMLRV 196
 154 APOLAKLQARSVSVKVGAYGIEDHLRQAFSTHLLVGGNPPATSTYTLIHALRENGV 213
 197 VFDLLKLDVATLFTSFTSKYFSQDKLRQVTSFTLLVGGNPLSPALYAMIHFEVKTWGI 256
 214 WFPRGGTGALVNGMIKLFODLGGBWLNARVSHMETVGD-----KIQAVOLEDGRFEIC 268
 257 HYAMGGTGAIVGLVQKEEELGGAIRGAGVDEVLVDGNLPKRTAGVRLSEGEELRAD 316
 269 AVASNADVVHTYRDLLOHQAAPAAQKAKLQSKRMN--SLFYLYFLGNHHD--QLAHTTV 325
 317 LVASNGDWANTY--LKRVRPSARLVNSDLRVKAASEMSLLVYVYFGRGDDLPKLNHNI 374

326 CFQPRYRELTHIEIFNHDGLARDFSLYHAPCVTDPSLAPEGCGSVYVLAPVPHLGTANED 385
 375 LLGPRYELLSEIFGTRKLGEDFSQYLHVPTITDALAPAGHAAATLVVPING-SGID 433
 386 WAVEGPRLRDRIPDYLEQH-YMFGRLRSQLVTHMTPTDFRDELNAWQGSASFVPILTQ 444
 434 WDVEGPKLAALADITERRGLIPGLRERLTHFEFTPTDYFAGTLDLYLGNAFGPPRLVQ 493
 445 SAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLLED 490
 494 SAFFRPHNRSEDLNHFYLVGAGQFGAGTPTSVMSAKMTARLAIED 539

RESULT 12
 Q8KZ28 PRELIMINARY; PRT; 514 AA.
 AC Q8KZ28;
 DT 01-OCT-2002 (TREMELrel. 22, Created)
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DB Phytoene dehydrogenase.
 GN CRTI.
 OS uncultured proteobacterium.
 OC Bacteria; Proteobacteria; environmental samples.
 OX NCBI_TaxID=153809;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21822632; PubMed=11832943;
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
 RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;
 RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
 RL Nature 415:630-633(2002).
 DR EMBL; AE008921; AAM48646.1; -.
 DR GO; GO:0018491; F:oxidoreductase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:006118; P:electron transport; IEA.
 DR InterPro; IPR002937; Amino oxidase.
 DR InterPro; IPR008150; Bac phytoene_dh.
 DR InterPro; IPR000205; NAD_B5.
 DR InterPro; IPR008151; Phytin dehydro.
 DR Pfam; PF01593; Amino oxidase; 1.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR PROSITE; PS00982; PHYTOENE DH; 1.
 SQ SEQUENCE 514 AA; 56824 MW; D5177500CAB56773 CRC64;

Query Match 37.8%; Score 984.5; DB 2; Length 514;
 Best Local Similarity 41.8%; Pred. No. 2.4e-71;
 Matches 206; Conservative 77; Mismatches 191; Indels 19; Gaps 5;

6 VIGAGFGLALAIRLQAGIPVLLLEQDKPGGRAYVYQ-EGQFTFDAGPTVITDPSAIE 65
 15 VIGAGLGLAAMRLGAKGYAVTVLDKLDVGGSGSVTQDGHFIDLGTIVTPKVFES 74
 66 LPALAGKQKDYVELLPVTPYRLCWESKGVNTNDQALEAQIQFNPRDVAGYRAFL 125
 75 LWAACGRDFHADVDLRPLEPFYBIRWPDGYSFRASGDDERKQSEVQRINPADLPGRFL 134
 126 DYSAVFNQGYLKLGTVPFLSKMQLRAAPQLAKLQARSVSVKVGIVDEHLRQAFSP 185
 135 KDSQKRYIIGEGVWAPFMHRLWTLKVLPTFAMLRADRSIYGLAARVVKDERLMAISP 194
 186 HSLVGGNPPATSTYTLIHALRENGVWFPRTGTGALVNGMTKLFODLGGVVLNARVS 245
 195 HPLFTGGDPHVTSTYALVAHLEKTYGVHVAGVQVQIADAMAADVRAQGGQI----- 247
 246 HMETVGDKI-----QAVOLEDGRFEETCAVASNADVVHTYRDLLOSHPAKQAKLIQ 298
 248 HQNAVADILLINGAQAVALTQGRFDADPLTVSNADAGHTYDHLNHSRRKRTTKLA 307
 299 SKRMSNLSFLYFLGNHHDQLA---EHTVCFQPRYRELTHIEIFNHDGLAEDFSLYLHAP 355
 308 RKRWSMGLFVWYFQTRGTAGRWADVGHHTIANGPRYKGLLRDIFLKGRLSDMSLYIHRP 367

```
2Y 356 CTTDPSLAPGCGSYVVLAPVPHLGTAN-LDNAVEGPRRLDRIDFVLEQHYMPGLRSQLV 414
DB 368 SVTPSPVAPAGDDTFYVLSVPHLGTANLDMAVEGPRRLDRIDFVLEQHYMPGLRSQLV 426
QY 415 THRMFTPTDFRDLNANWQGSFAFVPEPILTQSANFRPHNRDHDNDNLVLCAGTHPGAGIP 474
DB 427 TETFTPTDFRDLNANWQGSFAFVPEPILTQSANFRPHNRDHDNDNLVLCAGTHPGAGIP 486
QY 475 GVIGSAKATAGIM 487
DB 487 GVIGSAEVLKLV 499

RESULT 13
292G56 PRELIMINARY; PRT; 517 AA.
AC Q92GE4;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Diapophytoene dehydrogenase Crtn.
SN CRN.
SS Helicobacillus mobilis.
SC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
SC Helicobacillus.
XX NCBI_TaxID=28064;
XX [1]
XX SEQUENCE FROM N.A.
XX MEDLINE=93061957; PubMed=9843979;
XX Xiong J., Inoue K., Bauer C.E.;
XX "Tracking molecular evolution of photosynthesis by characterization of
XX a major photosynthesis gene cluster from Helicobacillus mobilis.";
XX Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
XX EMBL; AF080002; AAC84034.1; -.
XX PIR; T31463; T31463.
XX GO; GO:0016491; 2-oxido-reductase activity; IEA.
XX GO; GO:0006118; 2-oxido-reductase activity; IEA.
XX InterPro; IPR000759; Adrnx reductase.
XX InterPro; IPR002937; Amino oxidase.
XX InterPro; IPR008151; Phytan dehydro.
XX InterPro; IPR001100; Pyr redox.
XX Pfam; PF01593; Amino oxidase; 1.
XX PRINTS; PR00419; ADXRDTASE.
XX PRINTS; PR00411; ENDRDTASE.
XX ProDom; PD139017; Phytan dehydro; 1.
SQ SEQUENCE 517 AA; 58340 MW; BB933977999C4587 CRC64;

Query Match 28.9%; Score 752; DB 2; Length 517;
Best Local Similarity 34.2%; Pred. No. 2e-52;
Matches 170; Conservative 91; Mismatches 214; Indels 22; Gaps 7;

2Y 6 VIGAGFGGLALAIRLQAGIPVLLLEQRDKPGGRAYVYQGGFTPDAGPTVITDPS 65
DB 13 VVGSGAGGSAVLANQGDVTLREKATPGRLSAIQAGYSIDVGPTIMMNDVPHQ 72
QY 66 LPALAGKQKDYVELLPVTPFPYRLCHWESKGVNY--DNDQALQEAQIQFNPVAGRAFL 125
DB 73 YFKDGNIEDYLDVLRNFCYHLHFTDGTGCMKPSIDLKELDEIRSFNPDVDGGLYL 132
QY 126 -----DYSRAVNEGYLKLGTVPFLSPK-DMLRAAPQAKLQAVRSVYKAVGIEDE 177
DB 133 AQIHRRVQAREKIE---KSFTKPSDFNIDTLGNWQLRTL---NNVYDDIARFIKD 186
QY 178 HLRAQFPHLLVCGNPPATSSITLI-----HALREWGVPFRPGGTGALVNGIMKLFOD 233
DB 187 RLRLSLTFQAIYLVGSPDPAPSYTLIGVVEHGLS---GVWPKGGNAITQALVKLGE 243
QY 234 LGGGWLINARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVHTYRDLSSQHPAAQ 293
DB 244 FGGSLRVNAREVEQILIEQGRVAVGLRANGVELKADVVISNADFPYTWENLVPAHSGKXT 303
QY 294 AKKIQSKRMNSLFLVYFGLNHHHDQLAHHTVCGPRYRELIIHEIFNHDGLAEDFSLYLH 353
```

```
DB 304 POKLNSMEVTCGAFMLYLGVRNRYDNLHLNHYFTPDYKYSMDLFTROQLPQDFAMVY 363
QY 354 APCVTDLSLAPGCGSYVVLAPVPHLGTANLDMAVEGPRRLDRIDFVLEQHYMPGLRSQLV 413
DB 364 RPTKYDSDVAPGKDIIVLVVPENL--SSGIDWKKEHRYRELIVKGLERQGVTDLSKHI 422
QY 414 VTHRMFTPTDFRDLNANWQGSFAFVPEPILTQSANFRPHNRDHDNDNLVLCAGTHPGAGI 473
DB 423 EFERIYTPETFTQNRFTYQGAAGLAPSLFQSGYFERPHIKSEVPNLYFSGASVHPGGV 482
QY 474 PGVIGSAKATAGIMLED 490
DB 483 PVVLVCGKLVSEQVLAD 499

RESULT 14
Q9ENMO PRELIMINARY; PRT; 511 AA.
AC Q9ENMO;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase (phytone desaturase) (EC 1.3.-.-).
GN OB2461.
OS Oceanobacillus thevensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HT831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus thevensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004601; BAC14417.1; -.
DR GO; GO:0016491; 2-oxido-reductase activity; IEA.
DR GO; GO:0006118; 2-oxido-reductase activity; IEA.
DR InterPro; IPR000759; Adrnx reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR000205; NAD B5.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00411; ENDRDTASE.
XX Oxidoreductase; Complete proteome.
SQ SEQUENCE 511 AA; 58463 MW; BE70FD9F71D406AB CRC64;

Query Match 27.1%; Score 706.5; DB 16; Length 511;
Best Local Similarity 33.2%; Pred. No. 9.6e-49;
Matches 170; Conservative 91; Mismatches 212; Indels 39; Gaps 12;

QY 2 KPTTVIGAGFGGLALAIRLQAGIPVLLLEQRDKPGGRAYVYQGGFTPDAGPTVITDPS 61
DB 4 KKVIVIGAGVAGLALAIRLQAGFNVEYKESMPGKMHQITKGTGFTDGLSPVIMMPE 63
QY 62 AIELFALAGKQKDYVELLPVTPFPYRLCHWESKGVNY--DNDQALQEAQIQFNPVAGRAFL 119
DB 64 LYRAVVELTGRNPDYIPMERIDPMNVPFGNTPEYQISSDLIQLIDIEIAISKDAA 123
QY 120 GYRAFLD--YSRAVNEGYLKLGTVPFLSPK-DMLRAAPQAKLQ-----AMRSVYSKVA 171
DB 124 GFITVYLNKIYOR-----FQIAKDHFIQRFPRHPQWQFTPTKLWNLGLKRTL 169
QY 172 G-----YIEDHLRQAFSFLSLVGNPPATSSITLIHALREWGVPFRPGGTGAL 223
DB 170 GNADNFTKVKYKIDRSLRQMSFTQLYIGISPYNGPSLYTIIPMEIYLVGWYFKGGMYTM 229
QY 224 VNGMKLFQDLGSEWLINARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVHTYRDL 283
DB 230 ATSMKFLFELGGTIHNSVPEKVIIEKKATGKLEN-LEINADYVVCNADFPYSKML 288
QY 284 LSQHPAAAKQA-KKIQSKRMNSLFLVYFGLNHHHDQLAH-HTVCGPRYRELIIHEIFNH 341
```

289 IQDKAKGKYDDEKIDKMKYSCSCFIMYLGMNKYDEVSNNVHNPFFSENKMTMDIFTG 348
342 DGLAEFSELYLHAPCVTPSPSLAPGCGSYVYLAAPVPHLGTANLDMAVEG-PRLRDRIFDY 400
349 XRL-EDAFYTYIGSKWDPSLAPGKGIYVLPVSDLSAQYSGWDETIAAYREKVEK 407
401 LEQ-HYMPGLASQIVTHMFTFPFRDELNAWQGSFVEPILTQSAWFRPHNRDKHIDN 459
408 LDE-DBFKGIDEVISETYMTPFLDFESKFAYNACFGRLPTLSQSNHLRPSQKNCEN 467
460 LYLVGAGTHPCAGIPGVIGSAK-ATAGLMLEDL 490
468 LYFGSSTHPGAGVPIVLLSAKIATELLKDD 499

SULT 15
UF87
Q7UF87 PRELIMINARY; PRT; 537 AA.
01-OCT-2003 (TRENBLrel. 25, Created)
01-OCT-2003 (TRENBLrel. 25, Last sequence update)
01-OCT-2003 (TRENBLrel. 25, Last annotation update)
Phytoene dehydrogenase (EC 1.14.99.-).
CRTI OR RB10370.
Rhodospirillum baltica.
Bacteria; Planctomycetes; Planctomycetacia; Planctomycetiales;
Planctomycetaceae; Firellula.
NCBI_TaxID=117;
(1)
SEQUENCE FROM N.A.
STRAIN=1;
MEDLINE=22735913; PubMed=12835416;
Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
Schlesner H., Amann R., Reinhardt R.;
"Complete genome sequence of the marine planctomycete Firellula sp.
strain 1.";
Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
EMBL: BX294151; CAD78796.1; -;
Oxidoreductase; Complete proteome.
SEQUENCE 537 AA; 60136 MW; A16C389C746949B6 CRC64;

Query Match 26.6%; Score 693.5; DB 16; Length 537;
Best Local Similarity 32.5%; Pred. No. 1.2e-47;
Matches 162; Conservative 115; Mismatches 204; Indels 17; Gaps 9;
2 KPTTVIGAGFCGLALRLQAGIPVLLLRQKPGGRAYVYQGGFTPDAGTPTIDPS 61
25 KKVWVGAGPGGLASAKQLAAGCGDVTILERRGQVGRTSAIEIDGFRDCGPTFFLYR 84
62 AIELFALAGKQLDVYVELLPVTPYRLCWESGVFNVDQALQIAOQFNPRDVAGY 121
85 VLDRIFHSTGRDLMEVPMBELDPQYRLTGGGQQLDCTPDMEMBRQIAQFSQDVQQL 144
122 RAFLDYSAVFNKGLVTVPLFKMDLRAAPOL---AK-LQAMRSVYKVGAYIEDE 177
145 KRYMDNR-IXLEKRPILSPHPSALDVMK--PSLGAAGKHLHPFTLCKELERYFSOP 201
178 HLQAFSPHSLVCGNPFATSIYTLIALERWGVWFRGGTCAVNGMIKLPQDLGGR 237
202 RLVTAFQSKYLGNSPNCPSLSILSFLEYEGVPHPIGGCSRVSEKNAETAEEMGVK 261
238 VVLNARVSHMETVGDKIOAQVLEDRGRFETCAVASNADVVVHTYRDLSCH-PAAAKQA-- 294
262 IRLNEPVDSEMEGRVRALHTQ-ADRYDADA FVWNAD---FADWMTKTVPNASRKWS 316
295 -KKLQSKMSNSLFLVYGLNHHDLQAHVTCGPRYRLIHEI FNHDLAEDFSLYLH 353
317 DEQAKKKFSCSTWLYIGIEGLYDHPHSIHSSIDYNNRLREIETDHTLSQDPSVYQ 376
354 APCVTDPSLAPEGCGSYVYLAAPVPHLGTANLDMAVEGPRLRDRIFDYLEQHYMPGLRSOL 413

Db 377 NAGVTDPTLAPAGHSSLYLVVPVTH-DTENVDWKSAAAGPRELTLDKLGELGLTDVRDRI 435
QY 414 VTHRMFTFPFRDELNAWQGSFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPCAGI 473
Db 436 RVEHQITPDDHQSDYSYKGGATFNLAHNLGQMLHCRPRNRPFEELDGYLVGSGGTHPSSEL 495
QY 474 PGVIGSAKATAGLMLEDL 491
Db 496 PVIVESSRISRLLIQDL 513

Search completed: February 29, 2004, 14:51:01
Job time : 43.4556 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

* protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 8.29213 Seconds
(without alignments)
3089.496 Million cell updates/sec

File: US-09-941-947A-32

Effect score: 2606
Sequence: 1 MKPTVIGAGGGLAAILR.....IPGVIGSAKATAGLMLEDLI 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2438	93.6	492	1 CRTI_PANAN	P21685 pantoaea ana
2	2023	77.6	492	1 CRTI_ERWHE	P22871 erwinia her
3	1638	62.9	501	1 CRTI_AGRAU	P54978 agrobacteri
4	1007.5	38.7	524	1 CRTI_RHOCA	P17054 rhodobacter
5	959.5	36.8	518	1 CRTI_RHOSH	P54980 rhodobacter
6	757.5	29.1	517	1 CRTJ_MYXXA	P54979 myxococcus
7	672.5	25.8	507	1 CRTI_STRGR	P54961 streptomyce
8	660.5	25.3	508	1 CRTI_STRGR	P54971 streptomyce
9	635	24.4	595	1 CRTI_NEUCR	P21334 neurospora
10	616	23.6	529	1 CRTI_MYXXA	Q02861 myxococcus
11	595.5	22.9	583	1 CRTI_PHYBL	P54982 phycococcy
12	530	20.3	621	1 CRTI_CERNL	P48537 cercospora
13	430.5	16.5	495	1 CRTD_RHOSH	Q01671 rhodobacter
14	369.5	14.2	494	1 CRTD_RHOCA	P17059 rhodobacter
15	171.5	6.6	469	1 P49_STRLI	P06108 streptomyce
16	140	5.4	527	1 AOFB_HUMAN	P21397 homo sapien
17	140	5.4	749	1 TR2M_AGRHR	Q09109 agrobacteri
18	136.5	5.2	570	1 CRTI_SOYBN	P28553 glycine max
19	132.5	5.1	527	1 PROX_PROF	Q32434 propionibac
20	127.5	4.9	583	1 CRTI_LYCES	P28554 lycopersico
21	124.5	4.8	495	1 AOFN_ASPNG	P46882 aspergillus
22	123.5	4.7	516	1 OKLA_AGRGH	P81382 agkistrodon
23	121.5	4.7	571	1 CRTI_MALZE	P49086 zeaxanthin
24	121	4.6	527	1 AOFB_CANFA	P50027 canis fami
25	120.5	4.6	454	1 AOFB_MYCTU	Q53320 mycobacteri
26	116	4.5	526	1 AOFB_RAT	P21398 rattus norv
27	116	4.5	657	1 VH09_RALSO	Q8XYP9 ralstonia s
28	115	4.4	508	1 FMSI_YEAST	P50264 saccharomyc
29	114.5	4.4	557	1 TR2M_PSRSS	P06617 pseudomonas
30	113.5	4.4	566	1 CRTI_ORYSA	Q92709 oryza sativ
31	113	4.3	582	1 CRTI_CAPAN	P80093 capsicum an
32	112	4.3	527	1 AOFB_BOVIN	P21398 bos taurus
33	111.5	4.3	417	1 YAAB_RHISN	P55349 rhizobium s

34	110	4.2	519	1 AOFB_RAT	P19643 rattus norv
35	110	4.2	567	1 FIG1_HUMAN	Q961q9 homo sapien
36	110	4.2	999	1 OXRP_RAT	Q63617 rattus norv
37	109.5	4.2	444	1 YGAF_COLLI	P37339 escherichia
38	109.5	4.2	535	1 Y897_MYCTU	Q10555 mycobacteri
39	108.5	4.2	566	1 CRTI_ARATH	Q07356 arabidopsis
40	108	4.1	519	1 AOFB_HUMAN	P27338 homo sapien
41	107.5	4.1	471	1 PPOX_MYXXA	P56601 myxococcus
42	106	4.1	519	1 AOFB_CANFA	P58028 cavia porce
43	106	4.1	755	1 TR2N_AGRVI	P25017 agrobacteri
44	105.5	4.0	999	1 OXRP_HUMAN	Q9V411 homo sapien
45	104.5	4.0	999	1 OXRP_CRIGR	Q60432 cricetus

ALIGNMENTS

RESULT 1

ID	CRTI_PANAN	STANDARD;	PRT;	492 AA.
AC	P21685;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DB	Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).			
GN	CRTI..			
OS	Pantoaea ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoaea.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
RX	MEDLINE=9107214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway			
RT	by functional analysis of gene products expressed in Escherichia			
RT	Coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
CC	!- FUNCTION: This enzyme converts phytoene into lycopene via the			
CC	intermediaries of phytofluene, zeta-carotene and neurosporene by			
CC	the introduction of four double bonds.			
CC	!- COPACTOR: FAD (Probable).			
CC	!- PATHWAY: Carotenoid biosynthesis			
CC	!- SIMILARITY: Belongs to the phytoene dehydrogenase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; D90087; BAA14127.1; ..			
DR	PIR; D37802; D37802.			
DR	InterPro; IPR002937; Amino oxidase.			
DR	InterPro; IPR008150; Bac phytoene dh.			
DR	InterPro; IPR008151; Phytan dehydro.			
DR	ProDom; P01593; Amino oxidase; 1.			
DR	ProDom; P0139017; Phytan dehydro; 1.			
DR	PROSITE; PS00982; PHYTOENE DH; 1.			
KW	Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.			
FT	NP BIND 5 38 FAD (ADP PART) (POTENTIAL).			
SQ	SEQUENCE 492 AA; 55007 MW; P44CD4E34A9C6413 CRC64;			

Query Match 93.6%; Score 2438; DB 1; Length 492;
Best Local Similarity 93.5%; Pred. No. 3.8e-190;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKPTVIGAGGGLAAILRQAAGIPVLLLEQRDKPGRAYVYQEQGTFDAGPTVITDP 60
|||||

1 MKPTTVIGAGGGGLAALRAQLQAAGIPVLLLEQDKFGGRAYVYEDQGFDFDAGPTVITDP 60
61 SAIEELPALAGKOLKDYVELLPVTPFVRLCWSSGKVFYNDQOALQLEAGIQOQFNPRDVAG 120
61 SAIEELPALAGKOLKDYVELLPVTPFVRLCWSSGKVFYNDQOALQLEAGIQOQFNPRDVAG 120
121 YRAFLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKLQAWRSYYSKVAGYIIDEHLR 180
121 YRAFLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKLQAWRSYYSKVAGYIIDEHLR 180
181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMKLFOOLGGEVVL 240
181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMKLFOOLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKGLASK 300
241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKGLASK 300
301 RMNSLFLVYFGLNHHQDLAHTVCGPRYRELIIHIFNHDGLAEDFSLYLHAPCVITDP 360
301 RMNSLFLVYFGLNHHQDLAHTVCGPRYRELIIHIFNHDGLAEDFSLYLHAPCVITDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPRDRIFDYLEOHYMPGLRSOLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPRDRIFDYLEOHYMPGLRSOLVTHRMFT 420
421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 2

RTI_ERWHE
D_CRTI_ERWHE STANDARD; PRT; 492 AA.
C P22871;
T 01-AUG-1991 (Rel. 19, Created)
T 01-AUG-1991 (Rel. 19, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CRTI.
S Erwinia herbicola.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=549;
X [1]
P SEQUENCE FROM N.A.
C STRAIN=RH010;
X MEDLINE=91088634; PubMed=2263648;
X Armstrong G.A., Alberti M., Hearst J.E.;
T "Conserved enzymes mediate the early reactions of carotenoid
T biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";
X Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979 (1990).
X -!- FUNCTION: This enzyme converts phytoene into lycopene via the
X intermediaries of phytofluene, zeta-carotene and neurosporene by
X the introduction of four double bonds.
X -!- COFACTOR: FAD (Probable).
X -!- PATHWAY: Carotenoid biosynthesis.
X -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
X This SWISS-PROT entry is copyright. It is produced through a collaboration
X between the Swiss Institute of Bioinformatics and the EMBL outstation
X at the European Bioinformatics Institute. There are no restrictions on its
X use by non-profit institutions as long as its content is in no way
X modified and this statement is not removed. Usage by and for commercial
X entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
X or send an email to license@isb-sib.ch).
C R EMBL; M38423; AAA24820.1; -;
R EMBL; M87280; AAA64981.1; -;

DR A39273; A33120.
DR InterPro; IPR02937; Amino oxidase.
DR InterPro; IPR008150; Bac.phytoene dh.
DR InterPro; IPR008151; Phyto_dehydro.
DR Pfam; PF01593; Amino_oxidase; 1.
DR PRODOM; PD139017; Phyto dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 5 38 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8EDC5DB1562083F2 CRC64;
Query Match 77.6%; Score 2023; DB 1; Length 492;
Best Local Similarity 76.8%; Pred. No. 1.8e-156;
Matches 377; Conservative 42; Mismatches 72; Indels 0; Gaps 0;
QY 1 MKPTTVIGAGGGGLAALRAQLQAAGIPVLLLEQDKFGGRAYVYEDQGFDFDAGPTVITDP 60
DB 1 MKPTTVIGAGGGGLAALRAQLQAAGIPVLLLEQDKFGGRAYVYEDQGFDFDAGPTVITDP 60
QY 61 SAIEELPALAGKOLKDYVELLPVTPFVRLCWSSGKVFYNDQOALQLEAGIQOQFNPRDVAG 120
DB 61 SAIEELPALAGKOLKDYVELLPVTPFVRLCWSSGKVFYNDQOALQLEAGIQOQFNPRDVAG 120
QY 121 YRAFLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKLQAWRSYYSKVAGYIIDEHLR 180
DB 121 YRAFLDYSRAVFNQGYLKLGTVPFLSFKDMLRAAPQALAKLQAWRSYYSKVAGYIIDEHLR 180
QY 181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMKLFOOLGGEVVL 240
DB 181 QAFSFLHLLVGGNPPATSSITLILHALERWGVNPPGGTGALVNGMKLFOOLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKGLASK 300
DB 241 NARVSHMETVGDKIQAQVLEDEGRFETCAVASNADVVHTYRDLSSQHPAAQAQKGLASK 300
QY 301 RMNSLFLVYFGLNHHQDLAHTVCGPRYRELIIHIFNHDGLAEDFSLYLHAPCVITDP 360
DB 301 RMNSLFLVYFGLNHHQDLAHTVCGPRYRELIIHIFNHDGLAEDFSLYLHAPCVITDP 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPRDRIFDYLEOHYMPGLRSOLVTHRMFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGRPRDRIFDYLEOHYMPGLRSOLVTHRMFT 420
QY 421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDLNNAWQSAFSVEPIITQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
QY 481 KATAGLMLEDLI 491
DB 481 KATAGLMLEDLI 491

RESULT 3

CRTI_AGRAU
ID CRTI_AGRAU STANDARD; PRT; 501 AA.
AC P54978;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=44155;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96062243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,
RA Ohtani T., Miki W.;
RT "Structure and functional analysis of a marine bacterial carotenoid
RT biosynthesis gene cluster and astaxanthin biosynthetic pathway
RT proposed at the gene level.";

J. Bacteriol. 177:6575-6584 (1995).

-!- FUNCTION: This enzyme converts phytoene into lycopene via the intermediaries of phytylflavene, zeta-carotene and neurosporene by the introduction of four double bonds (By similarity).

-!- COFACTOR: FAD (Probable).

-!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin biosynthetic pathway.

-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; D58420; BAA09594.1; -
 InterPro; IPR002937; Amino oxidase.
 InterPro; IPR008150; Bac phytoene dh.
 Pfam; PF01593; Amino oxidase; 1.
 ProDom; PD139017; Phytin dehydro; 1.
 PROSITE; PS00982; PHYTOENE DH; 1.
 Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 SEQUENCE 501 AA; 54806 MW; 5P251AF11D679358 CRC64;

Query Match 62.9%; Score 1638; DB 1; Length 501;
 Best Local Similarity 63.9%; Pred. No. 3.3e-125;
 Matches 313; Conservative 61; Mismatches 116; Indels 0; Gaps 0;

2 KPTTVIGAGFGGLALAIQLQAGIPVILLLEQRKPGGRAYVYQGGTFDAGPTVITDPS 61
 5 KTAIVIGAGFGGLALAIQLQAGIATTLVEARDKPGGRAYVYVHDDGHLFDAGPTVITD 68
 62 ALEELFALAGKQLKDYVELLPVTFYRLCWSEKGVFNFDQGLENIQIQFNPDRVAGY 121
 69 ALXELWALTGDMDARDVTLMPVSPFYRLMWPGGKVFVDVNEADQLERQIAQFNPDDLEG 128
 122 RAFLDYSRAVFNFGYKLGTVFPFLSFKMDLRAAPQLAKLQAWRSVYSKVAGYETDEHLRQ 181
 129 RFRDIYAEVYQEGYKLGTVFPFLKLGQMLKAAPALMKLEAYKSVHAKVATFVKDPIRQ 198
 182 AFSFSLVGNPFTATSIYTLIHALERMGWFPFGGTGALVNGXIKLPDILGGEVVLN 241
 189 AFSYHTLLVGNPFTSIYALNHALERGGWFAKGTQNLVAGWALFERLGQMLN 248
 242 RVSWMETVGDKIQAQVLEDGRRFETCAVSNADVHTYRDLLSQHPAAAKQAKKLSKR 301
 249 AKVARIDTGRATGVTIADGRALTADNVASNGDVMNRYDRLIGHTARGOSRAKSLNAKR 308
 302 MSNSLFVLYFGLNHHHDLAHTVCFGRYRELHIEIPNHDGLAEDPSLYLHAPCVITDPS 361
 309 WMSLFLVLFHGLREAPKDVAAHTILFGRYKELVNEIFKPKLAEDPSLYLHSPCTDPE 368
 362 LAPEGCGSYVYLAIPHGLTANLWAVEGPRILDRIDYLEQHYMPGLRQLVTHRMFT 421
 369 MAPPGMSTYVYLAIPHGLGRADIDWAVEGPRYADRIILASLEERILPNLANLTTRIFTP 428
 422 PDRDELNAWGSAPSVPIITQSAWFRPHNRDGHIDNLYLVAGTHPGAGIPQVIGSAK 481
 429 SDFASELNAHGSAPSVPIITQSAWFRPHNRDXTIRNFVLVAGTHPGAGIPQVIGSAK 488
 482 ATAGLMLEDL 491
 489 ATAQWMLSDL 498

SULT 4

TI RHOC
 CRTI RHOC
 P17054;

PRT; 524 AA.

STANDARD;

Query Match 38.7%; Score 1007.5; DB 1; Length 524;
 Best Local Similarity 41.5%; Pred. No. 5.1e-74;
 Matches 204; Conservative 89; Mismatches 194; Indels 5; Gaps 3;

DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 GN CRTI.
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 CX NCBI_TaxID=1061;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=SB1003 / St Louis, and BEC404;
 RX MEDLINE=89313663; PubMed=2747617;
 RA Armstrong G.A., Alberti M., Leach F., Hearst J.E.;
 RT "Nucleotide sequence, organization, and nature of the protein
 products of the carotenoid biosynthesis gene cluster of Rhodobacter
 capsulatus.";
 RT Mol. Gen. Genet. 216:254-268 (1989).
 RL [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=89327279; PubMed=2546948;
 RA Bartley G.E., Scolnik P.A.;
 RT "Carotenoid biosynthesis in photosynthetic bacteria. Genetic
 characterization of the Rhodobacter capsulatus CrtI protein.";
 RL J. Biol. Chem. 264:13109-13113 (1989).
 RN [3]
 RN ERRATUM.
 RA Bartley G.E., Scolnik P.A.;
 RL J. Biol. Chem. 264:18260-18260 (1989).
 RN [4]
 RN SIMILARITY TO CAROTENOID DESATURASES.
 RX MEDLINE=90368827; PubMed=2144293;
 RA Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
 RT "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
 crassa are structurally and functionally conserved and contain
 domains homologous to flavoprotein disulfide oxidoreductases.";
 RL J. Biol. Chem. 265:16020-16024 (1990).
 CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
 intermediary of phytylflavene by the symmetrical introduction of two
 double bonds at the C-11 and C-11' positions of phytoene.
 CC -!- COFACTOR: FAD (Probable).
 CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; J04969; AAA50313.1; -
 EMBL; X52291; CAA36533.1; -
 EMBL; Z11165; CAA77540.1; -
 FIR; A32617; A32617.
 InterPro; IPR000759; Admrx_reductase.
 InterPro; IPR002937; Amino oxidase.
 InterPro; IPR008150; Bac phytoene dh.
 InterPro; IPR000205; NAD BS.
 InterPro; IPR008151; Phytin dehydro.
 Pfam; PF01593; Amino oxidase; 1.
 PRINTS; PR00419; ADXEDTASE.
 ProDom; PD139017; Phytin dehydro; 1.
 PROSITE; PS00982; PHYTOENE DH; 1.
 Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 Oxidoreductase; FAD; Flavoprotein; NAD.
 NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 SEQUENCE 524 AA; 57978 MW; 6425A7ESA06A6B9 CRC64;


```

422 EDRDE-LANAGCGAFSVEPILTSQAMFRPNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
434 ETRDYLSPW-GAGFSTIEPILOSAFRPNISSEVANLFLYGAGTHPGAGIPGVIGSA 492
481 KATAGL 486
493 EWMAKL 498

RESULT 6
TO MYXXA
CRUI MYXXA STANDARD; PRT; 517 AA.
P54979;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase).
CARA2.
Mycococcus xanthus.
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Cystobacterineae; Myxococcaceae; Myxococcus.
NCBI_TaxID=34;
[1]
SEQUENCE FROM N.A.
STRAIN=DK1050;
MEDLINE=96061955; PubMed=7588751;
Botella J.A., Ruiz-Vazquez R.M.;
"A cluster of structural and regulatory genes for light-induced
carotenogenesis in Myxococcus xanthus.";
Eur. J. Biochem. 233:238-248(1995).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- FACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene desaturase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z21955; CAA79956.1; -
PIR; S32169; S32169.
InterPro; IPR000759; Adrndx_reductase.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR000205; NAD_BS.
InterPro; IPR008151; Phytin dehydro.
Pfam; PF01593; Amino oxidase; 1.
PRINTS; PR00419; ADXRDTASE.
ProDom; PD139017; Phytin dehydro; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 11 44 FAD (ADP PART) (POTENTIAL).
SEQUENCE 517 AA; 57783 MW; CECDD74AB9F9F8CB CRC64;

Query Match 29.1%; Score 757.5; DB 1; Length 517;
Best Local Similarity 35.6%; Pred. No. 9.8e-54;
Matches 173; Conservative 88; Mismatches 216; Indels 9; Gaps 6;

6 VVGAGGGLAALAILOAGIPVLLLRDKPKGGRAYVYQGGTFPAGPTVTDPSAISE 65
12 VVGAGVGGLAALAAALHAGQGFVQVFEKTPQGGRCNRLQVDGFTWDLGPTVILMPVFEE 71
66 LFALAGKOLXDYVELLPVTPYRLCMESGVENVYDNDQALQIQGFNPRDVGAYRAFL 125
72 TFRVAGRIEDYILLKCDPNRVHFRDRSDVFTSILCMAGRELREVEGSGVARYLAF 131
126 DYSAVFNEGYKLGTVFFLSFKDML--RAAPLAKLQAWRSVYSKVAGYIEDEHLRQAF 183

```

```

Db 132 AGRVQVYETSLDHLVGNRYAGLRDYLSPRVLARLQVRAHRMTADVERFFQDERLRAM 191
Qy 184 SFHSLVGNPNFATSSYITLHALERBNGVWPPRGGTGALVNGMKLQFDLGGGVNLNAR 243
Db 192 TFOYWLGVSPVSPAVYGLLPFTLGLGVWFPKGLVAIPQALERLARERBGRVPHYCAP 251
Qy 244 VSEMETVGDKIOAVOLEDCGRPFETCANASNDVHTYRDILLSQHPAAAKQAKLOSKEMS 303
Db 252 VERILTDGGRTRGRVLEGGVEADAVLCNADLVAYEKLDPKATTLKRKEKL---RYT 308
Qy 304 NSLFVLYEGLNHHDLQLAHVTVCFGPRYRELIIHFNHDLGLAEDFSLYHAPCVTDPSLA 363
Db 309 SSGYMLYLGKRRYPPELLHNVVFGRDYKGSFDDIPEF-RVPEDPSFYVNAFTETDASLA 367
Qy 364 PEGCGSYVILAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMGLASQL-VTHRMETPF 422
Db 368 PEGKDALYVLVPEFH-QEPDLDMKVEGPKVRAKFAFARVABELGFPFLESDEVERRSSTPD 426
Qy 423 DFRDELNAWQGSFSAFSEVPELITQSAWFRPNRDKHIDNLYLVGAGTHPGAGIPGVIGSAK- 481
Db 427 DWAGTFNLARGSGFOLSONFIQIGFFPFSNQDARVKNLFFVGASTQPGTGLPTVLISARL 486
Qy 482 ATAGLM 487
Db 487 VTERLM 492

RESULT 7
CRUI STRGR STANDARD; PRT; 507 AA.
AC P54981; P72447;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI OR CRTPE.
OS Streptomyces griseus.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JA3933;
RX MEDLINE=97074881; PubMed=9917308;
RA Schumann G., Nurnberger H., Sandmann G., Kruegel H.J.;
RT "Activation and analysis of cryptic crt genes for carotenoid
biosynthesis from Streptomyces griseus.";
Mol. Gen. Genet. 252:658-666(1996).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
CC -!- FACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene desaturase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; L37405; AAA91950.1; -
EMBL; X95596; CAA64850.1; -
DR InterPro; IPR000759; Adrndx_reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac_phytoene_dh.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.

```

R ProDom; PD139017; Phytin dehydro; 1.
 R PROSITE; PS00982; PHYTOENE DH; 1.
 W Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 T NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 Q SEQUENCE 507 AA; 5450 MW; FBB57F7FE696B2AC CRC64;

Query Match 25.8%; Score 672.5; DB 1; Length 507;
 Best Local Similarity 35.0%; Pred. No. 7.6e-47;
 Matches 171; Conservative 80; Mismatches 214; Indels 10; Gaps 10;

Y 6 VIGAGGGLALATLQAAGIPVLLLBQRDKPGGRVYVQEGFTFDAGTWTIDPSAIEE 65
 b 13 VVGAGLAGLAAHLHLCAGRRVTVVEREDVPGGRAGLLEGGFRIDTGTVTIMPLDVED 72
 Y 66 LFALAGKQLKDYVELLPVTPFYELCWESKVFNYNDQALEAQIQFNPDRVAGYRFL 125
 b 73 AFAAVGERMADRLRLAPAYRARPADGSLDVTHTDGAAMEAAVEEF-----AGARQAV 127
 Y 126 DYSR-AVFREGYKLGTVPL--SPKMLRAA-POLAKLQA---WRSVYKVGAGYIEDEH 178
 b 128 GYRLRLIWLRLYRVMRRPDTNFDSPQLAHPDLARLAALGGFRDLARIGHFVSDER 187
 Y 179 LRQAQFPHSLLVGNGNPFATSSITLHALERENGWVFPFGGTGALVNGMKLFDQLGGEV 238
 b 188 LRVPFSQALYAGVPPARALAAVAVIAYMDTVAGVYFPRGGMHALPRAMADAASDAGAVL 247
 Y 239 VLNRVSHMETVGGKCAVQLEGGRRPFCPCAVASNADVVHTTRDILLSQHPAAAKKQLQ 298
 b 248 RYGGRTVRLERSGDRVTAV-VTDQRIADAVVLTDPDLPVSYR-LIGRTP-----HRPL 299
 Y 299 SKMNSLFLVYFGLNHHHDQLAHTVCFGRYRELHIEIFNHGDLAEFSLYLHAPCVT 358
 b 300 PLRHSFSAVILHGTDTWDLAHTTISFGAAWNTFHELTTRGLMSDPSLLITRPTAT 359
 Y 359 DPSLAPEGCSYVYLAPEVPH--LGTANLDWAVEGPRRLDRIDFYLEQHYNPGRLSOLVTH 416
 b 360 DPSLAPGKHLHYLAPCPNTEVGPGRVRELGPRYRDELLAELAEEREMPGGLAAIEE 419
 Y 417 RMTPEPDRDELNAWGSASFVEPILTQSAFWPFRPHNRKHIDNLXLVGAGTHPGAGIPCV 476
 b 420 GLVTPVDWNTAQGHA-AGTPPSVAHTPTQGPFRPNLRVGTVNAVLAGCGTTPGVGVFTV 478
 Y 477 IGSAKATA 484
 b 479 LISGKLA 486

RESULT 8

RTI STRSE STANDARD; PRT; 508 AA.
 ID CRTI STRSE
 AC P54971;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 CR CRTI.
 SN Streptomyces setonii.
 SC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 SC Streptomycineae; Streptomycetaceae; Streptomyces.
 SC NCBI_TaxID=38315;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ISP 5395;
 RA Hoshi K.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 SC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the intermediary of phytofluene by the symmetrical introduction of two double bonds at the C-11 and C-11' positions of phytoene.
 SC -!- COFACTOR: FAD (Probable).
 SC -!- PATHWAY: Carotenoid biosynthesis.
 SC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
 SC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; D55723; BAA09537.1; -
 DR InterPro; IPR000759; Admrx_reductase.
 DR InterPro; IPR002937; Amino oxidase.
 DR InterPro; IPR008150; Bac_phytoene_dh.
 DR InterPro; IPR000205; NAD_BS.
 DR InterPro; IPR008151; Phytin_dehydro.
 DR Pfam; PF01593; Amino oxidase; 1.
 DR PRINTS; PR00419; ADXRDASE.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR PROSITE; PS00982; PHYTOENE DH; 1.
 KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
 FT NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 508 AA; 54610 MW; 4BDEFC076D51CB5 CRC64;

Query Match 25.3%; Score 660.5; DB 1; Length 508;
 Best Local Similarity 34.7%; Pred. No. 7.2e-46;
 Matches 173; Conservative 69; Mismatches 213; Indels 43; Gaps 11;

QY 6 VIGAGGGLALATLQAAGIPVLLLBQRDKPGGRVYVQEGFTFDAGTWTIDPSAIEE 65
 Db 13 VVGAGLAGLAAHLHLCAGRRVTVVEREDVPGGRAGLLEGGFRIDTGTVTIMPLDVED 72
 QY 66 LFALAGKQLKDYVELLPVTPFYELCWESKVFNYNDQALEAQIQFNPDRVAGYRFL 121
 Db 73 AFAAVGERMADRLRLAPAYRARPADGSLDVTHTDGAAMEAAVEQFAGARQAVGYRFL 132
 QY 122 -----RAFLDYSAVFNEGYKLGTVPF--LSPKMLRAAPOLAKLQAWESVYS 168
 Db 133 RIMLERLYRVQMRFDID---ANFDS-----PFOLVHPDLAR-----LAALGGFRELDA 177
 QY 169 KVAGYIIDEHLRQAFSPHSLLVGNGNPFATSSITLHALERENGWVFPFGGTGALVNGMI 228
 Db 178 RIGHFVSDERLRLVRFSPQALYAGVPPARALAAVAVIAYMDTVAGVYFPRGGMHALPRAMA 237
 QY 229 KLFDLGGVTVLNRVSHMETVGGKCAVQLEGGRRPFCPCAVASNADVVHTTRDILLSQHP 288
 Db 238 DAADAGASFRYQGSVTRLERSGDRVTAV-VTDQRIADAVVLTDPDLPVSYR-LIGRSP 295
 QY 289 AAAQAQKLOSKMNSLFLVYFGLNHHHDQLAHTVCFGRYRELHIEIFNHGDLAEF 348
 Db 296 -----HRLPLRHSFSAVILHAGTDTWDLAHTTISFGAAWNTFHELTTRGTGLMSDP 349
 QY 349 SLYLHAPCVTDPDPSLAPEGCSYVYLAPEVPH--LGTANLDWAVEGPRRLDRIDFYLEQHYM 406
 Db 350 SLLITRPTASDPSLAPEGSKHLHYLAPCPNTEVGPGRVRELGPRYRDELLAELAEEREM 409
 QY 407 PGLRSQVTHRMFTPEPDRDELNAWGSASFVEPILTQSAFWPFRPHNRKHIDNLXLVGAG 466
 Db 410 PGLGSAIEVGLVTPVDWNTAQGHA-AGTPPSVAHTPTQGPFRPNLRVGTVNAVLAGCG 468

467 THPGAGIPGVIGSAKATA 484
 469 TTPGVGVPTVLISGKLA 486

RESULT 9

CRTI NEUCR STANDARD; PRT; 595 AA.
 ID CRTI NEUCR
 AC P21334;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase) (Albino-1 protein).
 GN AL-1.
 OS Neurospora crassa.


```

QY 118 VAGYRAZLDYSGRAVNBGLKL--GTVPFLSPKDMLEBAAPQAKLQAWRSVYKVGAGYIE 175
DB 128 ASALRQWVEDGREKYIAYQKICTSADNIGYAPWELAPTL-RFPWQTYLQDGFTH 186
QY 176 DEHLRQAFSFLSLVGGNPFATSIYTLIHALEREWGVPFGGTGALVNGMIKLPQDLG 235
DB 187 DDRVTYALAYPSKYLGHLHTCCSVFSVFPFELAFGVHVEGFGFRLSRGWRKARDLG 246
QY 236 GEVVLNARVSHMETVGDKIQAQVLEDGRRETCAVASNADVVHTYRDLLSHQHAARAKQAK 295
DB 247 ATRFMGTPEVKRVADGAVGKLVGGVEVDADAVVNAADLAYAARSLSI---PARAREGS 303
QY 296 K-----LQSKMNSLFLVYFGLNHHHDQLAHTVCFGRPYRLIHIFNHDGL----- 344
DB 304 RLDDAALERAKYSCSTFWAYGLDVTYADLPPLHLYLSESART----DRDALEDRHVD 358
QY 345 AEDSLYIHACVTDPSLAPGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLEQH 404
DB 359 LEDPPFYVNCVGTDPGAPAGHSTLYLVPTNTGRP--VDWVKTBOALRERIPAMLEKV 417
QY 405 YMPGLRSOLVTHRMFTPFDFRDELNAWOGSAFSEPILTQSAMFRPHNRDKHIDNLYLVG 464
DB 418 GLKGVREHIREERTFTAEHRDDFNVFRAVENLSHTWLQGLPLRPVKVNRDIEGLYFVG 477
QY 465 AGTHPGAGIPGVIGSAKATAGIMLED 490
DB 478 GGTHPGSGLLTIMESANIAADYLRE 503

RESULT 11
CRTL_PHYBL STANDARD; PRT; 583 AA.
AC P54982;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
EN CARB.
OS Phycomyces blakesleeanus.
DC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
DC Phycomyces.
DX NCBI_TaxID=4837;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1555;
RA Ruiz-Hidalgo M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X78434; CAA55197.1; -.
DR PIR; S43139; S43139.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR000205; NAD ES.
DR Pfam; PF01593; PhytN dehydro.
DR ProDom; PD139017; PhytN dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;

```

```

KW Transmembrane.
FT NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
FT TRANSMEM 531 551 POTENTIAL.
SQ SEQUENCE 583 AA; 65983 MW; 808F682B12FB591 CRC64;

Query Match 22.9%; Score 595.5; DB 1; Length 583;
Best Local Similarity 29.1%; Pred. No. 1.6e-40;
Matches 146; Conservative 103; Mismatches 225; Indels 27; Gaps 12;

QY 2 KPTTVIGAGFGSLALAIQLQAAGIPVLLLEQRDKEGGRVAVYQEQGTFPDAGPTVITDPS 61
DB 5 KXVIIIGAGAGTATAARLARBEGIKVTVEKNPGGRCSLINHNGRFPQGSFLYMPK 64
QY 62 AIEELFALAGKQDYVELLPVTFYFLCHWESGKVFNVDNQAOLEAQIOOFNPRDVAGY 121
DB 65 LFEEAFEALDEKIDHVELLRCHNNYKVFHDDGDKIQIQLSSDLGRMKPEMERIRGPD--GF 122
QY 122 RAFLDYSPA--VNEGYLKLGTVPFLSKDM-----LRAAPQLAKLQAWRSVYKVGAGYI 174
DB 123 LRPLDFMKESTHYEGGVEWAIKQ--NPFETIWLKILRLQYVPAFLRLHIFDFVYSRAAKYF 180
QY 175 BDEHLRQAFSFLSLVGGNPFATSIYTLIHALEREWGVPFGGTGALVNGMIKLPQD 233
DB 181 KTKMREMAFTQSMYMGMSPYDSPAVYNELQYTFEASGIWVPGGFNTVIQKLENIATEK 240
QY 234 LGGEVVLNARVSHMET--VGDKIQAQVLEDGRRETCAVASNADVVHTYRDLLSHQHPAAA 291
DB 241 FGARFIYEAPVAKINTDDKGGKVTGVTLQSGEVIADAVVNAADLAYAARNLL---PPCR 297
QY 292 KQAKCLQSKRMSNLSFLVYFGLNHHHDQLAHTVCFGRPYRLIHIFNHDGLAEDPSLY 351
DB 298 WTTNTLAELKLTSSISFYNSLKEVPSLDVHNLFLAEAFKESFDEIFTDBKMPSELSFY 357
QY 352 LHAPCVTDPSLAPGCGSYVVLAPVPHLGTAN---LDWAVEGPRLEDRIFDYLEQHMPG 408
DB 358 VNLFSRIDPTAAPGKDSMIVLPIGHMKSKTNEADYTMIVKEARQVLEVLRL--G 415
QY 409 LRS--QLVTH-RMFTPFDFRDELNAWOGSAFSEPILTQSAMFRPHNRDK--HIDNLYLV 463
DB 416 LTFNFDLVEHEEVNDPSINQKFKLWRGSLGLSHDVLQVLMFRPSTQDSTGYKMLFFV 475
QY 464 GAGTHPGAGIPGVIGSAKATA 484
DB 476 GASTHPGTGVPVLVLAGSKLTS 496

RESULT 12
CRTL_CERN
ID CRTL_CERN STANDARD; PRT; 621 AA.
AC P48537;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN PDH1.
OS Cercospora nicotianae.
DC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
DC Dothideomycetes et Chaetochyomycetes incertae sedis;
DC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cercospora.
DX NCBI_TaxID=29003;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 18366;
RC MEDLINE=94368091; PubMed=8085820;
RA Ehrenschaft M., Daub M.E.;
RT "Isolation, sequence, and characterization of the Cercospora
RT nicotianae phytoene dehydrogenase gene";
RL Appl. Environ. Microbiol. 60:2766-2771(1994).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.

```


-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

C	EWBL; U03903; AAB86988.1; --			
R	PIR; T48646; T48648.			
R	InterPro; IPR008150; Bac_phytoene dh.			
R	InterPro; IPR008151; Phyt_n dehydro.			
R	ProDom; PD139017; Phyt_n dehydro; 2.			
R	PROSITE; PS00982; PHYTOENE DH; 1.			
R	Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;			
W	Transmembrane.			
W	NP BIND 11 44 FAD (ADP PART) (POTENTIAL).			
T	TRANSMEM 536 536 POTENTIAL.			
Q	SEQUENCE 621 AA; 6199EA7784963CAB CRC64;			

Query Match	20.3%;	Score 530;	DB 1;	Length 621;
Best Local Similarity	26.1%;	Pred. No. 3.7e-35;		
Matches 145;	Conservative 90;	Mismatches 224;	Indels 96;	Gaps 11;

[illegible]

```

120  y  GYRAFLDYGRAVFNCGYLKLGTPVPLSPKD--MLRAPO---LAKIQAWRSVYKQVAG 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125  b  GYTRYLEPLKEGSHGVYELSVREVLINTEGLTAMLR--PEPIRLLQLHPFESINIRAGK 182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
173  y  YIEDHLRQAFSFHLLVCGNPFATSSYTIHIALERENGWVFPRGCTGALVNGMKLQFO 232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183  b  YFWTERLRVRFEGSMVGMSPFDAPGYSYLLQVTELAEGIWIYVGGFHVAVCALVKIGE 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

233	y	DLGGEVLNARVSHMETVGDKI-----QAVQLEDGRFFETCAVSNADVVVHYRDL	283
243	b	REGVDF-----RMEATVKILLSEDDGVAKGVELEDGRRLADVVVNSDLVAYEKL	295
284	y	LSCHPAAAKQAKKLOSKMSNSFLVYFGLNHHHDQLAHTVCGPYRRLIHEIENNHDG	343
296	b	L---PIKTPVASSLKGKRGSCSSISFYWALOROVPELEAHNIPLADEYRSSFPSFISFKKHL	352

```

344 LAEDFSYLHAPCVTDPSLAPEGQSGYYVLAPVPHL----- 379
      ::::
      ::::
353 IPDPSFYVNPSPRVDSTAAPGKDSVVVLVPVGHLLLEDRHASQAQLSASRNHSSA 412
      ::::
      ::::
380 -----GTANLDWAVEGPRLRDIPDYLEQHYMPCRLSOLVTHRMFTFPDFRDLNA 430
      ::::
      ::::
413 SPDPDGLPTEKODWPAMISLARKTILSTIOSNTVDVLTPIIITHESTNSPI.SWKOTFNI 472
      ::::
      ::::

```

```

431 WOGSAFSEVPILTQSAWERP-----HNRDKGIDN 459
      |||
473 DRGATGLSHSEFNVLCFRPTTRAKPGAFDAQLLKFGVLGRAEVIIDAFRCRGDKG 532
      |||
460 LYLVGAGTHSGAGIP 474
      |||
533 LYMVGASAFHGGTGV 547
      |||

```

RESULT 13	
RTD_RHOSH	STANDARD;
Q01671	495 AA.
01-JUL-1993	(Rel. 26. Created)
16-OCT-2001	(Rel. 40. Last sequence update)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

28-FEB-2003 (Rel. 41, Last annotation update)
Methoxynuospirore dehydrogenase (EC 1.14.99.-).
GN
CNID.
Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
CC
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC
Rhodobacteraceae; Rhodobacter.
NCBI_TaxID=1063;
NCBI_TaxID=1063;
CC

LI
NP SEQUENCE FROM N.A., / 2.4.1 / NCIB 8253 / DSM 158;
NC STRAIN=ATCC 17023 /
NX MEDLINE=9230736; Pubmed=162413;
PA Gari S.; Toledo J.C., Gilbert I., Barbe J.;
PM Nucleotide sequence of the methylenesulfoxide dehydrogenase gene
PT from Rhodobacter sphaeroides: comparison with other bacterial
PI carotenoid dehydrogenases. // 198(1993)

RA SEQUENCE FROM N.A.
 RM STAIN-ATCC 7203² / 2, 4, 1 / NCIB 8253 / DSM 158;
 RX MEDLINE-95378228; PubMed-7721699;
 RA Lang H.F., Cogdell R.J., Takachi S., Hunter C.N.;
 RI "Complete DNA sequence, specific rms, insertion map, and gene
 DI assignment of the carotenoid biosynthesis pathway of Rhodospirillum rubrum." *J. Biol. Chem.* 267:10681-10686 (1992).

RL spharotates, 177-2064-2073 [1995].
 CC U. Bacteriol.
 CC 1- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR
 CC 2- NEOSPHEROIDENE.
 CC 3- METABOLITE: NEOSPHEROIDENE.
 CC 4- CATABOLITE: FAD (Probable).
 CC 5- PATHWAY: Carotenoid and Chlorophyll biosynthesis.
 CC 6- PATHWAY: Carotenoid and Chlorophyll biosynthesis.
 CC 7- SIMILARITY: Belongs to the phytoene desaturase family.
 CC 8- SOURCE: *Neurospora crassa*.
 CC 9- SOURCE: *Neurospora crassa*.
 CC 10- SOURCE: *Neurospora crassa*.
 CC 11- SOURCE: *Neurospora crassa*.
 CC 12- SOURCE: *Neurospora crassa*.
 CC 13- SOURCE: *Neurospora crassa*.
 CC 14- SOURCE: *Neurospora crassa*.
 CC 15- SOURCE: *Neurospora crassa*.
 CC 16- SOURCE: *Neurospora crassa*.
 CC 17- SOURCE: *Neurospora crassa*.
 CC 18- SOURCE: *Neurospora crassa*.
 CC 19- SOURCE: *Neurospora crassa*.
 CC 20- SOURCE: *Neurospora crassa*.
 CC 21- SOURCE: *Neurospora crassa*.
 CC 22- SOURCE: *Neurospora crassa*.
 CC 23- SOURCE: *Neurospora crassa*.
 CC 24- SOURCE: *Neurospora crassa*.
 CC 25- SOURCE: *Neurospora crassa*.
 CC 26- SOURCE: *Neurospora crassa*.
 CC 27- SOURCE: *Neurospora crassa*.
 CC 28- SOURCE: *Neurospora crassa*.
 CC 29- SOURCE: *Neurospora crassa*.
 CC 30- SOURCE: *Neurospora crassa*.
 CC 31- SOURCE: *Neurospora crassa*.
 CC 32- SOURCE: *Neurospora crassa*.
 CC 33- SOURCE: *Neurospora crassa*.
 CC 34- SOURCE: *Neurospora crassa*.
 CC 35- SOURCE: *Neurospora crassa*.
 CC 36- SOURCE: *Neurospora crassa*.
 CC 37- SOURCE: *Neurospora crassa*.
 CC 38- SOURCE: *Neurospora crassa*.
 CC 39- SOURCE: *Neurospora crassa*.
 CC 40- SOURCE: *Neurospora crassa*.
 CC 41- SOURCE: *Neurospora crassa*.
 CC 42- SOURCE: *Neurospora crassa*.
 CC 43- SOURCE: *Neurospora crassa*.
 CC 44- SOURCE: *Neurospora crassa*.
 CC 45- SOURCE: *Neurospora crassa*.
 CC 46- SOURCE: *Neurospora crassa*.
 CC 47- SOURCE: *Neurospora crassa*.
 CC 48- SOURCE: *Neurospora crassa*.
 CC 49- SOURCE: *Neurospora crassa*.
 CC 50- SOURCE: *Neurospora crassa*.
 CC 51- SOURCE: *Neurospora crassa*.
 CC 52- SOURCE: *Neurospora crassa*.
 CC 53- SOURCE: *Neurospora crassa*.
 CC 54- SOURCE: *Neurospora crassa*.
 CC 55- SOURCE: *Neurospora crassa*.
 CC 56- SOURCE: *Neurospora crassa*.
 CC 57- SOURCE: *Neurospora crassa*.
 CC 58- SOURCE: *Neurospora crassa*.
 CC 59- SOURCE: *Neurospora crassa*.
 CC 60- SOURCE: *Neurospora crassa*.
 CC 61- SOURCE: *Neurospora crassa*.
 CC 62- SOURCE: *Neurospora crassa*.
 CC 63- SOURCE: *Neurospora crassa*.
 CC 64- SOURCE: *Neurospora crassa*.
 CC 65- SOURCE: *Neurospora crassa*.
 CC 66- SOURCE: *Neurospora crassa*.
 CC 67- SOURCE: *Neurospora crassa*.
 CC 68- SOURCE: *Neurospora crassa*.
 CC 69- SOURCE: *Neurospora crassa*.
 CC 70- SOURCE: *Neurospora crassa*.
 CC 71- SOURCE: *Neurospora crassa*.
 CC 72- SOURCE: *Neurospora crassa*.
 CC 73- SOURCE: *Neurospora crassa*.
 CC 74- SOURCE: *Neurospora crassa*.
 CC 75- SOURCE: *Neurospora crassa*.
 CC 76- SOURCE: *Neurospora crassa*.
 CC 77- SOURCE: *Neurospora crassa*.
 CC 78- SOURCE: *Neurospora crassa*.
 CC 79- SOURCE: *Neurospora crassa*.
 CC 80- SOURCE: *Neurospora crassa*.
 CC 81- SOURCE: *Neurospora crassa*.
 CC 82- SOURCE: *Neurospora crassa*.
 CC 83- SOURCE: *Neurospora crassa*.
 CC 84- SOURCE: *Neurospora crassa*.
 CC 85- SOURCE: *Neurospora crassa*.
 CC 86- SOURCE: *Neurospora crassa*.
 CC 87- SOURCE: *Neurospora crassa*.
 CC 88- SOURCE: *Neurospora crassa*.
 CC 89- SOURCE: *Neurospora crassa*.
 CC 90- SOURCE: *Neurospora crassa*.
 CC 91- SOURCE: *Neurospora crassa*.
 CC 92- SOURCE: *Neurospora crassa*.
 CC 93- SOURCE: *Neurospora crassa*.
 CC 94- SOURCE: *Neurospora crassa*.
 CC 95- SOURCE: *Neurospora crassa*.
 CC 96- SOURCE: *Neurospora crassa*.
 CC 97- SOURCE: *Neurospora crassa*.
 CC 98- SOURCE: *Neurospora crassa*.
 CC 99- SOURCE: *Neurospora crassa*.
 CC 100- SOURCE: *Neurospora crassa*.

CC This SwissProt entry is copyrighted, it is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC in the European Bioinformatics Institute, there are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

[illegible]

KW	Photosynthesis;	Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW	Oxidoreductase;	FAD; Flavoprotein; NAD;
NP_BIND	9	42 (ADP PART) (POTENTIAL).
FT	CONFLICT	A -> G (IN REF. 2).
FT	CONFLICT	T -> P (IN REF. 2).
FT	CONFLICT	L -> M (IN REF. 2).
FT	CONFLICT	L -> I (IN REF. 2).
FT	CONFLICT	PEGA -> ATPG (IN REF. 1).
FT	CONFLICT	

SQ SEQUENCE 495 AA; 52900 NM; 62227931415B353E CRC64;
 Query Match 16.5%; Score 430.5; DB 1; Length 495;
 Best Local Similarity 29.4%; Pred. No. 3.2e-27;
 Matches 150: Conservative 64; Mismatches 243; Indels 53; Gaps 15;

6	VIGAFGGGLALAIRLQAAGIPVLLLEQORDKPGRAVVYQEQFTFDAGTVTITDPSAIE	65
		::
10	VVGAWGGGLASAIRLARAGEVTLLLEAREAPGGRMRTLPVAPGPDAGTVTITLREVPD	69
		::
66	LFALAGKQIKDVLLPTTPPVRLCWESGKVFNYND-QAOLEAGIQOFPNPDVAGYAP	124
		::
70	IFEVCGKIDHHTLLPQPLARHNLPGSTLDTLTDLEANVZAAVAPAGAREAPAFRRF	129
		::

```

125 LDYSAVFNEGYLKLTVPFLSF-KDMLRAA-POL-----AKLOAWRSVYSKVAG----- 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 HDLSARLYD-----AFDREMNRAARPDURAIATGALKAPRTWPALLPGMTLDR 177

```


This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Y00142; CAA68336.1; -;
PIR; S09189; S09189.
InterPro; IPR000759; Adrndx_reductase.
InterPro; IPR000205; NAD_BS-
InterPro; IPR008152; Phytin_dehydro.
PRINTS; PRO0419; ADXRDYASE.
ProDom; PD139017; Phytin_dehydro; 1.
SEQUENCE 469 AA; 49740 MW; C2AE7533C7C701CB CRC64;

Query Match 6.6%; Score 171.5; DB 1; Length 469;
Best Local Similarity 23.3%; Pred No. 3.1e-06;
Matches 126; Conservative 64; Mismatches 212; Indels 139; Gaps 29;
1 MKPTTVICAGFGGLALAIRLOAGIPVLLLEQRDKPGRAYVYQ--EQGFED----AGP 54
1 MLDVVVGAGPGLTAAVELARRGPPVAVPEAOQTVGGARTTELTLPGRHDPSCSAHP 60
55 TVITDPSAIEELPALAGKQKDY-VEL-----LPVT-PFYRLCWESGKVFNYDNDQOLEA 108
61 LGINSP-----APRGPLERYGLEWMLHPGLPMAHPF-----PDGSAAVLSR 101
109 QIQO-----FNPRDVAGYRAFLDYGRAVFNEGYKLGTVPLSFKDM-LRAAPQ----LAK 159
102 SVGETAASFQDAGAPYRLI-----ERFLPRWDTLARDFMSLELTALPRDPVTILAR 153
160 L-----QAWRSVYKVGAGVIEDEHLRQAPS-----FHSLL-VGGNPPATSSIVTLIHAL- 207
154 FGLVGLPEPTW-----LMRRFRDEKAKTLFAGLVAVHMAPLGG--FATGAI-GLVFALA 204
208 --EREWGVWFRGGTGAIVNGMIKLFQDLGGEVVLNARVSHMETVGDKIQAVOLEDGRFP 265
205 AHAGWVY--ARGSQSISDALTYLKDGLGAVHTDYEVKRLD-----DLPPARAY 253
266 ETCVASNADVHTVTRDLSQHPAAAKQAKLQSK-----RMSNSLFVLYFGLNHHHDQLA 321
254 -----VLDTSPTALARIAGLGSYHANYRGSPVKIDYALDGPVPWTA 296
322 HH-----TVCFGPYREL---IHEIFNHDGLAEDFSLYLHAPCVTDPSLAPEGCGSYVYL 373
297 EEPKSAGTVQIGADSTEIGAAHAPSGTDRAPERPFITVQPSVADPTAPAGKHVFWAY 356
374 APVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVLSVTHRMFTPTDFRDELNAWQG 433
357 GHVENGWNGDLTDAME--QLER-----FAPGFRDVLARATAGPAELAARNANYVG 406
434 SAFP-----VEPILTOSAMFRPHNRDKHDNLYLVGAGTHPGAGIPGVIG--SAK 481
407 GDISSGAVSGQLLLRPKISLFPYSTPH-----PAVFICSSATPPGPGVHGMSGHNAAK 460
482 A 482
461 A 461

arch completed: February 29, 2004, 14:45:12
b time : 10.2921 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 13.6946 Seconds
(without alignments)
3455.835 Million cell updates/sec

file: US-09-941-947A-32

fect score: 2606

quence: 1 MKPTTVTGAGGGLAALRL.....IPGVIGSAKATAGLMLLELI 492

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tail number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

1: Pir1:**

2: Pir2:**

3: Pir3:**

4: Pir4:**

Pred. No. is the number of results predicted by chance to have a
score Greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	2453	94.1	492	2	S52586	phytoene dehydroge
2	2438	93.6	492	2	D37802	phytoene dehydroge
3	2023	77.6	492	2	A33120	phytoene dehydroge
4	1099.5	42.2	511	2	T50910	phytoene dehydroge
5	1051	40.3	548	2	C75466	phytoene dehydroge
6	1007.5	38.7	524	2	A32617	phytoene dehydroge
7	959.5	36.8	518	2	S49620	phytoene dehydroge
8	932.5	36.6	518	2	T50745	phytoene dehydroge
9	757.5	29.1	517	2	S32169	hypothetical prote
10	752	28.9	517	2	T31463	probable diaphophyt
11	635	24.4	595	2	A35919	carotenoid biosynt
12	629.5	24.2	523	2	T36968	probable phytoene
13	624	23.9	502	2	B90061	squalene synthase
14	616	23.6	529	2	S35306	phytoene dehydroge
15	604	23.2	499	2	AG2209	zeta-carotene desa
16	595.5	22.9	583	2	S43139	phytoene dehydroge
17	592	22.7	499	2	S43324	zeta-carotene desa
18	567	21.8	490	2	AH1199	phytoene dehydroge
19	562	21.6	582	2	T46822	phytoene desaturas
20	549.5	21.1	514	2	B69108	phytoene dehydroge
21	541.5	20.8	512	2	H84320	phytoene dehydroge
22	530	20.3	621	2	T48646	Phytoene dehydroge
23	493	18.9	448	2	B55348	crtn protein - Sta
24	489	18.8	488	2	AF1557	phytoene dehydroge
25	482	18.5	497	2	E90061	hypothetical prote
26	466	17.9	528	2	T51119	phytoene desaturas
27	433.5	16.6	536	2	B84327	Phytoene dehydroge
28	427.5	16.4	486	2	S49624	methoxyneurosporen
29	423.5	16.3	495	2	S23633	methoxyneurosporen

30	411	15.8	485	2	S32171	hydroxyneurosporen
31	388.5	14.9	486	2	T50749	methoxyneurosporen
32	380.5	14.6	542	2	S76617	hypothetical prote
33	375	14.4	565	2	A12273	hypothetical prote
34	369.5	14.2	494	2	S04406	methoxyneurosporen
35	343	13.2	454	2	A99470	phytoene dehydroge
36	342.5	13.1	498	2	H83880	hypothetical prote
37	322	12.4	506	2	AB2064	hypothetical prote
38	316	12.1	511	2	B75561	probable phytoene
39	313	12.0	525	2	T50893	methoxyneurosporen
40	311	11.9	501	2	S75951	hypothetical prote
41	292.5	11.2	587	2	AB6203	hypothetical prote
42	289	11.1	543	2	G87635	phytoene dehydroge
43	276.5	10.6	518	2	G90413	phytoene dehydroge
44	274.5	10.5	506	2	T34971	probable carotenoi
45	259	9.9	380	2	A12185	hypothetical prote

ALIGNMENTS

RESULT 1

S52586

phytoene dehydrogenase (EC 1.3.-.-) - Erwinia herbicola

C:Species: Erwinia herbicola

C>Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999

C:Accession: S52586

R:Lin, Y.P.; Lai, E.M.; To, K.X.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tni000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52586

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-492 <Lin>

A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21263.1; PID:G148397

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992

C:Superfamily: phytoene dehydrogenase

C:Keywords: oxidoreductase

Query Match	94.1%	Score	2453;	DB	2;	Length	492;
Best Local Similarity	93.7%	Pred. No.	5.9e-189;				
Matches	461;	Conservative	14;	Mismatches	17;	Indels	0;
Gaps	0;						
Qy	1	MKPTTVTGAGGGLAALRLQAGIPVLLLEQDKPKGRAYVYCEGFTDAGTWTDP	60				
Db	1	MKPTTVTGAGGGLAALRLQAGIPVLLLEQDKPKGRAYVYCEGFTDAGTWTDP	60				
Qy	61	SAIEELFALAGKQKQYVELLPVTPFYRLCWESGKVFYNDNDQALEAQIQQFNPRDVAG	120				
Db	61	SAIEELFTLAGKQKQYVELLPVAPFYRLCWESGKVFYNDNDQALEAQIQQFNPRDVEG	120				
Qy	121	YRAFLDYSRAVFNHGYKLGTVFPLSFQDMLRAAPOLAKLOAWRSYVSKVAGYIEDEHLR	180				
Db	121	YRQFLDYSRAVFNHGYKLGTVFPLSFQDMLRAAPOLAKLOAWRTYVSKVASYIEDEHLR	180				
Qy	181	QAFSFHLLVGGNPPFATSSITLIHALERNGVWFPFGGTGALVNGMKLFDGLGGEVVL	240				
Db	181	QAFSFHLLVGGNPPFATSSITLIHALERNGVWFPFGGTGALVNGMKLFDGLGGEVVL	240				
Qy	241	NARVSHMETVGDKIQAQVLEDGRRPFTCAVSNADVVHYTRDLLSOHPAAAKQKQSK	300				
Db	241	NARVSHMETVGDTEAVHLEDGRRPFTCAVSNADVVHYTRDLLSOHPAAVQSKQKQSK	300				
Qy	301	RMSNSLPVLYFGLNHHDDLAHTVCGFPYRLIHEIFNHDGLAEDFSLYLHAPCVTPD	360				
Db	301	RMSNSLPVLYFGLNHHDDLAHTVCGFPYRLIHEIFNHDGLAEDFSLYLHAPCVTDS	360				
Qy	361	SLAPEGCGSYVILAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQLVTHRMT	420				
Db	361	SLAPEGCGSYVILAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQLVTHRMT	420				
Qy	421	PFDFRDLNAAQSSAIVSEFILLTQSAFWFRPHNEDKHDNLVYLVGAGTHPCAGIPGVIGSA	480				

```
b 421 PFDFRQLNAYQSSAFSVEPVLTSQANFRPHNRDKTINNLVLVGAGTHPGAGIPGVGSA 480
y 481 KATAGLMLEDLI 492
b 481 KATAGLMLEDLI 492

RESULT 2
37802
hyoene dehydrogenase (EC 1.3.-.-) crti - Erwinia uredovora
;Species: Erwinia uredovora
;Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
;Accession: D37802
;Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashima, K.;
;Bacteriol. 172, 6704-6712, 1990
;Title: Blucidation of the Erwinia uredovora carotenoid biosynthetic pathway by function
;Reference number: A37802; MUID:91072214; PMID:2254247
;Accession: D37802
;Status: preliminary
;Molecule type: DNA
;Residues: 1-492 <MIS>
;Cross-references: GB:D90087; NID:g216681; PIDN:BAAL4127.1; PID:g216685
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoeductase

Query Match 93.6%; Score 2438; DB 2; Length 492;
Best Local Similarity 93.3%; Pred. No. 9,5e-188;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

y 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
b 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
y 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
b 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
y 121 YRAFLDYSRAVFNVEGYLKGTVPFLLSFKDMLRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
b 121 YRAFLDYSRAVFNVEGYLKGTVPFLLSFKDMLRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
y 181 QAFSFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 240
b 181 QAFSFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 240
y 241 NARVSEMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQHPAAKQKLOSK 300
b 241 NARVSEMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQHPAAKQKLOSK 300
y 301 RMNSLFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
b 301 RMNSLFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
y 361 SLAPEGCGSYVYLAPVPHLGTANLDNAVSGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
b 361 SLAPEGCGSYVYLAPVPHLGTANLDNAVSGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
y 421 PFDFRQLNAYQSSAFSVEPVLTSQANFRPHNRDKTINNLVLVGAGTHPGAGIPGVGSA 480
b 421 PFDFRQLNAYQSSAFSVEPVLTSQANFRPHNRDKTINNLVLVGAGTHPGAGIPGVGSA 480
y 481 KATAGLMLEDLI 492
b 481 KATAGLMLEDLI 492

RESULT 3
33120
hyoene dehydrogenase (EC 1.3.-.-) - Erwinia herbicola
;Alternate names: phytoene desaturase
;Species: Erwinia herbicola
;Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 29-Sep-1999
```

```
C;Accession: A39273; A33120
R;Armstrong, G.A.; Alberti, M.; Hearst, J.E.
Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990
A;Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nor
A;Reference number: A39273; MUID:9108634; PMID:2263648
A;Accession: A39273
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-492 <ARM>
A;Cross-references: GB:M38423; NID:gl48401; PIDN:AAA24820.1; PID:g148402
A;Note: the authors translated the codon CAG for residue 181 as Phe, TCC for residue 18'
A;Note: Strain Eho10; ATCC 39368
C;Genetics:
C;Superfamily: phytoene dehydrogenase
C;Keywords: carotenoid biosynthesis; oxidoeductase

Query Match 77.6%; Score 2023; DB 2; Length 492;
Best Local Similarity 76.8%; Pred. No. 2e-154;
Matches 377; Conservative 42; Mismatches 72; Indels 0; Gaps 0;

Qy 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
Db 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEQDKFGGRAYVYQGGTFDAGPTVITDP 60
Qy 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
Db 61 SAIEELPALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQALQAEQIQOQFNPRDVAG 120
Qy 121 YRAFLDYSRAVFNVEGYLKGTVPFLLSFKDMLRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
Db 121 YRAFLDYSRAVFNVEGYLKGTVPFLLSFKDMLRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
Qy 181 QAFSFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 240
Db 181 QAFSFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 240
Qy 241 NARVSEMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQHPAAKQKLOSK 300
Db 241 NARVSEMETVGDKIQAQVLEDGRFRFTCAVSNADVVHTYRDLSSHQHPAAKQKLOSK 300
Qy 301 RMNSLFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
Db 301 RMNSLFLVLYGGLNHHDDLAHTVCFGRPRYRELHEIFNHDGLAEDFSLYLHAPCVTDP 360
Qy 361 SLAPEGCGSYVYLAPVPHLGTANLDNAVSGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
Db 361 SLAPEGCGSYVYLAPVPHLGTANLDNAVSGPRLRDRIPDYLEQHYMPGLRSOLVTHRMFT 420
Qy 421 PFDFRQLNAYQSSAFSVEPVLTSQANFRPHNRDKTINNLVLVGAGTHPGAGIPGVGSA 480
Db 421 PFDFRQLNAYQSSAFSVEPVLTSQANFRPHNRDKTINNLVLVGAGTHPGAGIPGVGSA 480
Qy 481 KATAGLMLEDLI 491
Db 481 KATAGLMLEDLI 491

RESULT 4
T50910
phytoene dehydrogenase [imported] - Rubrivivax gelatinosus
C;Species: Rubrivivax gelatinosus
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
C;Accession: T50910
R;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynt
A;Reference number: Z25270
A;Accession: T50910
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-511 <NAG>
A;Cross-references: EMBL:AB034704; PIDN:BA94063.1
```

Experimental source: strain IL144
Genetics: crtl
Superfamily: phytoene dehydrogenase

Query Match 42.2%; Score 1099.5; DB 2; Length 511;
Best Local Similarity 42.3%; Pred. No. 3e-80;
Matches 204; Conservative 100; Mismatches 177; Indels 1; Gaps 1;
6 VTGAGGGLALAIQLAAGIPVLLLEQRDPKGGRAVYVQGGFTPDAGPTVTDPSAIBE 65
23 VWSGFGGMAAAVLAAGRYVTVLEKLDAPGGRAVYHRRGHVFDAGPTIVTPVPLFDE 82
66 LFALAGKOLADYVLELLPVTFFYRLCWESGVFNVDNDQAOLEACIQCPNPRDVAGYRAFL 125
83 LWALAGKRFSDDEILKSLDFYIRPDGDHPYSGDPAKRAEVRIRISPSDAEGFERFW 142
126 DYSRAVNEGYLKGTVTFPLSPKMDLRAAPQLAKLQWRSVSVKAGYIIEHLRQAFS 185
143 READQCYELGFTLGDRAFTVGLIKAAPLIVKLRGWSRLHOMVSHLKHPIKRIAMSL 202
186 HSLVAGNPNFATSIYLIHALEREVGFPRGGTGAIVNGMIKLFQDLGGEVVNLNARS 245
203 QSLGIGNPFSVSMYALVWALRGVHWAMGGTGELIRGLVDVFEFGMGTRRLKAEVK 262
246 HMETVGDKIQAQVLEDCRRFETCAVASNADVHTYRDLSSQHPAAKQAKLQSKMSNS 305
263 RIEVDNGVATGVTLADGERIPADIVVCGDTGYLYKXULVDARWKETDARIEGHYSMG 322
306 LFVLYFGLNHHDLAHTVCFGRPRRELHIEIFNHDGLAEDPSLYLHAPCVTDPSLAPE 365
323 LFVYFGTDRYEDVPHMVLGPRYRELDDIFRKKKLASDESIFYLHPTATDPSMAFA 382
366 GCGSYVYVLAIVPHLGTANLDWAVEGPRRLRIIDYLEQHYMGLRQLVTHRMFTPDFR 425
383 GCDTFYALMPVPHLGSCT-DWTQAEYRQSVQEALETIVLPGQLHRAVSFTCTPLDFQ 441
426 DELNMGGSFVSVEPIITQSAWFRPHNRKHIDNLYLVGAGTHPGAGIPGVIGSAKATAG 485
442 HRLLSYKAGFGLLEPLLQSAWFRPHNRSDVKNLFWVGASTHPGAGVPGVIMSAKALS 501
486 LM 487
502 VL 503
SULT 5
5466
ytoene dehydrogenase - Deinococcus radiodurans (strain R1)
Species: Deinococcus radiodurans
Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
Accession: C75466
White, O.; Eisen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
; Smith, H.O.; Venter, J.C.; Fraser, C.M.
ence 286, 1571-1577, 1999
Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
Reference number: A75250; MUID:20036896; PMID:10567266
Accession: C75466
Status: preliminary
Molecule type: DNA
Residues: 1-548 <WHI>
Cross-references: GB:AF001940; NID:G6458577; PIDN:AAF10439.1; PID:G645858
Experimental source: strain R1
Genetics:
Gene: DR0861
Superfamily: phytoene dehydrogenase
Map position: 1
Query Match 40.3%; Score 1051; DB 2; Length 548;
Best Local Similarity 43.5%; Pred. No. 2.6e-76;
Matches 229; Conservative 72; Mismatches 185; Indels 40; Gaps 9;

QY 2 KPTTVIGAGGGLALAIQLAAGIPVLLLEQRDPKGGRAVYVQ-BQGFTPDAGPTVITDP 60
DB 17 KTAIVIGAGGGLALAIQLAAGIPVLLLEQRDPKGGRAVYVQ-BQGFTPDAGPTVITDP 76
QY 61 SAIEELFA-----LAGKQKLD-----YVELLPVTPFYRLCWES 93
DB 77 HFIEELFALERDRAGLDAPDPYPPVLSGERVKGSGPHTSRVYVTLVPLPFIYRIVFHD 136
QY 94 GKFNVDNDQAOLEACIQCPNPRDVAGYRAFLDYSRAVNEGYLKGTVTFPLSPKMDLRA 153
DB 137 GTYFDYDGDPESTRQIAELAPGLAGYERPHADAEAFRRGFLGELGYTHFGDVPTMLRV 196
QY 154 APQAKLQWRSVSVKAGYIIEHLRQAFSFSHLLVGGNPPATSSITYLIHALEREVW 213
DB 197 VEDLLKDAVTLFSTFSKYFSQDKLRQVFSFETLLVGGNPLSVPAIYAMHFVEKTWGI 256
QY 214 WFPGGTGAIVNGMIKLFQDLGGEVVNLNARSVHMETVGD-----KIQAQVLEDCRRFETC 268
DB 257 HYAMGGTGAIVNGMIKLFQDLGGEVVNLNARSVHMETVGD-----KIQAQVLEDCRRFETC 268
QY 269 AVASNADVHTYRDLSSQHPAAKQAKLQSKMSN--SLFVLYFGLNHHHD-QLAHTTV 325
DB 317 LVASNGDWANTY--LKRVRFSARLVNSDLAVKASSEMSLLVYFGRGDDLPKXHHNI 374
QY 326 CFGPRYRELHIEIFNHDGLAEDPSLYLHAPCVTDPSLAPEGCGSYVYVLAIVPHLGTANLD 385
DB 375 LLGPRYELLASEIETGKRLGDEFSQYLFVPTLTDPALAPAGHAAAYTLVVPVPHNG-SGID 433
QY 386 WAVEGPRRLRIIDYLEQHYMGLRQLVTHRMFTPDFRDELNMGGSFVSVEPIITQ 444
DB 434 MDVGGPKLAELAAADIERRGLIPGLRERLTHFTPTDYPAGTLDLSYLGNAFGPEPRLVQ 493
QY 445 SAWFRPHNRKHIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLED 490
DB 494 SAFFRPHNRSEDHNFYLVGAGAGCPGAGTSPVMSAOWTARLIAED 539
RESULT 6
A32617
phytoene dehydrogenase (EC 1.3.-.-) - Rhodobacter capsulatus
N:Alternate names: phytoene desaturase
C:Species: Rhodobacter capsulatus
C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 29-Sep-1999
C:Accession: A32617; S04402
R:Bartley, G.E.; Scolnik, P.A.
J. Biol. Chem. 264, 13109-13113, 1989
A:Title: Carotenoid biosynthesis in photosynthetic bacteria. Genetic characterization of
A:Reference number: A32617; MUID:89327279; PMID:2546948
A:Accession: A32617
A:Molecule type: DNA
A:Residues: 1-524 <BAR>
A:Cross-references: GB:J04969; NID:G340554; PIDN:AAAS0313.1; PID:G556397
R:Armstrong, G.A.; Alberti, M.; Leach, F.; Hearst, J.E.
Mol. Gen. Genet. 216, 254-268, 1989
A:Title: Nucleotide sequence, organization, and nature of the protein products of the ca:
A:Reference number: S04401; MUID:89313663; PMID:2747617
A:Accession: S04402
A:Molecule type: DNA
A:Residues: 1-524 <ARM>
A:Cross-references: EMBL:X52291; NID:G45996; PIDN:CAA36533.1; PID:G45998
A:Note: translation of codons 1-33 is not given
A:Note: the authors translated the codon GTG for residue 34 as Met
C:Genetics:
C:Gene: crtI
C:Superfamily: phytoene dehydrogenase
C:Keywords: carotenoid biosynthesis; oxidoreductase
Query Match 38.7%; Score 1007.5; DB 2; Length 524;
Best Local Similarity 41.5%; Pred. No. 7.6e-73;
Matches 204; Conservative 89; Mismatches 194; Indels 5; Gaps 3;
QY 1 MKPTTVIGAGGGLALAIQLAAGIPVLLLEQRDPKGGRAVYVQ-BQGFTPDAGPTVITDP 60

```
b 8 MGRAVVIAGLGLAANRLGANGKYVTVVDRDLRDPGGRGSSITKGGHRFDLGPITVTP 67
y 61 SAIEELPALAGKQLKDYVELLPVTPPYRLCWESGKVYNDNDQAOLEAQIQFNPDRVAG 120
b 68 DRURELWADCGRDFDKVSLVPMPEYITDFDGEKYTAIGDADAKYAEVARIISPGDVEG 127
y 121 YRAFLDYSRAVNEGYKLGTVPFLLSFKDLRAAPQAKLQAWRSVYKVAQYIEDEHLR 180
b 128 FRHFMDAKARYEFGYENLGRKPMKLDLKVLPFGWLRADRSVYGHAKGVKDDHLR 187
y 181 QAFSFLSLVGNPPATSSITLIIHALEREWGVPFRGGTGALVNGMIKLFODLGGVVL 240
b 188 FALSFPFLPFGDPPFHTVSYILVSQLEKKFGVHAIGVQAIDAMAKVITDQGENRL 247
y 241 NARVSHMETVGD-KIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQS 299
b 248 NTEVDILVSRDGKATGIRLMOTELPAQVVVSNADAGHTYKRLLENDRDWTDEKLDK 307
y 300 KRMSNLSFLVLPGLN--HHHDLQAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPC 356
b 308 KRWSMGLFVYFGTKTAKMKVDVGHHTVVVGPVRYKEHVQDIFIKGELAEADMVYHRPS 367
y 357 YTDPSLAPEGCGSYVYLAAPVPHLGTAN-LDWAIVEGPRLDRIFDYLEQHYMPGLSOLVT 415
b 368 VTDPTAAPKGGDDTFYVLSPVNLDGPDGVDWSVEABKYAKVZKVIERLLFGVAEKITE 427
y 416 HRMFTDFDFRDELNAQGSFSAFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPG 475
b 428 EVVTFPTFEDRLVSLPGAGFSLEPILQSAWFRPHNASEEVDGLYLVGAGTHPGAGVPS 487
y 476 VIGSAKATAGLM 487
b 488 VIGSGSLVAQMI 499

RESULT 7
49620
hytoene dehydrogenase (EC 1.3.-.-) - Rhodobacter sphaeroides
;Alternate names: phytoene desaturase
;Species: Rhodobacter sphaeroides
;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 13-Sep-1998
;Accession: S49620
;Submitted to the EMBL Data Library, November 1994
;Description: The complete DNA sequence, specific TMS insertion map and gene assignment
;Reference number: S49619
;Accession: S49620
;Status: preliminary
;Molecule type: DNA
;Residues: 1-518 <LAN>
;Cross-references: EMBL:X82458; NID:G575405; PID:G575407
;Genetics:
;Gene: ctri
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoreductase

Query Match 36.8%; Score 959.5; DB 2; Length 518;
Best Local Similarity 40.9%; Pred. No. 5.3e-69;
Matches 199; Conservative 79; Mismatches 201; Indels 7; Gaps 5;

y 6 VIGAGFGLALAIRLQAAGIPVILLBQRDKPGGRAYVYQEGFTFDAGTIVTIDPSAIBE 65
b 15 VIGSGLGLAAANRLGAKGWRVTVIDKLDVPGGRGSSITQEGHRFDLGPITVTPQSLRD 74
y 66 LPALAGKQLKDYVELLPVTPPYRLCWESGKVYNDNDQAOLEAQIQFNPDRVAGVRAFL 125
b 75 LMTKCGRDFDADVELKPIDPFYEVWRPDSGSHFTVROSTAMKAEVARUSFGDVAGYEKFL 134
y 126 DYSAVNEGYKLGTVPFLLSFKDLRAAPQAKLQAWRSVYKVAQYIEDEHLQAFSP 185
b 135 KDEKRYWFGVEDLGRSMHKLWDLIKVLPTFGMWRADRSVYQHAALRVKDELRWALSF 194
y 186 HSLLVGNPPATSSITLIIHALEREWGVPFRGGTGALVNGMIKLFODLGGVVLNARS 245
b 188 FALSFPFLPFGDPPFHTVSYILVSQLEKKFGVHAIGVQAIDAMAKVITDQGENRL 247
y 241 NARVSHMETVGD-KIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQS 299
b 248 NTEVDILVSRDGKATGIRLMOTELPAQVVVSNADAGHTYKRLLENDRDWTDEKLDK 307
y 300 KRMSNLSFLVLPGLN--HHHDLQAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPC 356
b 308 KRWSMGLFVYFGTKTAKMKVDVGHHTVVVGPVRYKEHVQDIFIKGELAEADMVYHRPS 367
y 357 YTDPSLAPEGCGSYVYLAAPVPHLGTAN-LDWAIVEGPRLDRIFDYLEQHYMPGLSOLVT 415
b 368 VTDPTAAPKGGDDTFYVLSPVNLDGPDGVDWSVEABKYAKVZKVIERLLFGVAEKITE 427
y 416 HRMFTDFDFRDELNAQGSFSAFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPG 475
b 428 EVVTFPTFEDRLVSLPGAGFSLEPILQSAWFRPHNASEEVDGLYLVGAGTHPGAGVPS 487
y 476 VIGSAKATAGLM 487
b 488 VIGSGSLVAQMI 499
```

```
Db 195 HPLFIGDGPNTVSMYILVSQLEKEFGVHVAIGVAAIAAAMAKVIEGGGSRFMMTEVD 254
Qy 246 HMETVGDKIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSKEMNS 305
Db 255 EILVEKGTATGVLAGEVLRAGLVVSNADAGHTYKRLLENDRDWTDEKLDK 314
Qy 306 LFVLYFGLNHHH---DQLAHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVTDPSL 362
Db 315 LFVWYFGTKGTGMWPDVGHHITVNAAPRYKVLGVDEFILKGLKAKOMSLYIHRPSITDPTV 374
Qy 363 APGCGSYVYLAAPVPHLGTAN-LDWAIVEGPRLDRIFDYLEQHYMPGLSOLVTHRMETP 421
Db 375 APGDDDTFVALSPVPHLQKQAPVDVMOQVAEPYRESVLEVEQS-MPGIGERIGPSLVFTP 433
Qy 422 FDRDE-LNAQGSFSAFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPVIGSA 480
Db 434 ETRDRYLSW-GAGFSIEPILQSAWFRPHNISEVANLFLVAGTHPGAGVPGVIGSA 492
Qy 481 KATAGL 486
Db 493 EVNAKL 498

RESULT 8
T50745
phytoene dehydrogenase (EC 1.3.-.-) [imported] - Rhodobacter sphaeroides
;Species: Rhodobacter sphaeroides
;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 02-Sep-2000
;Accession: T50745
;R:Choudhary, M.; Kaplan, S.
;Nucleic Acids Res. 28, 862-867, 2000
;A>Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides ;
;Reference number: 225222; MUID:20115911; PMID:10648776
;Accession: T50745
;A>Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-518 <CHO>
;Cross-references: EMBL:AF195122; PIDN:AAF24289.1
;Experimental source: strain 2.4.1
;Genetics:
;Gene: ctri
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoreductase

Query Match 36.8%; Score 952.5; DB 2; Length 518;
Best Local Similarity 40.3%; Pred. No. 1.19e-68;
Matches 199; Conservative 78; Mismatches 202; Indels 7; Gaps 5;

Qy 6 VIGAGFGLALAIRLQAAGIPVILLBQRDKPGGRAYVYQEGFTFDAGTIVTIDPSAIBE 65
Db 15 VIGSGLGLAAANRLGAKGWRVTVIDKLDVPGGRGSSITQEGHRFDLGPITVTPQSLCD 74
Qy 66 LPALAGKQLKDYVELLPVTPPYRLCWESGKVYNDNDQAOLEAQIQFNPDRVAGVRAFL 125
Db 75 LMTKCGRDFDADVELKPIDPFYEVWRPDSGSHFTVROSTAMKAEVARUSFGDVAGYEKFL 134
Qy 126 DYSAVNEGYKLGTVPFLLSFKDLRAAPQAKLQAWRSVYKVAQYIEDEHLQAFSP 185
Db 135 KDEKRYWFGVEDLGRSMHKLWDLIKVLPTFGMWRADRSVYQHAALRVKDELRWALSF 194
Qy 186 HSLLVGNPPATSSITLIIHALEREWGVPFRGGTGALVNGMIKLFODLGGVVLNARS 245
Db 195 HPLFIGDGPNTVSMYILVSQLEKEFGVHVAIGVAAIAAAMAKVIEGGGSRFMMTEVD 254
Qy 246 HMETVGDKIQAOLEDDRRPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSKEMNS 305
Db 255 EILVEKGTATGVLAGEVLRAGLVVSNADAGHTYKRLLENDRDWTDEKLDK 314
Qy 306 LFVLYFGLNHHH---DQLAHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVTDPSL 362
Db 315 LFVWYFGTKGTGMWPDVGHHITVNAAPRYKVLGVDEFILKGLKAKOMSLYIHRPSITDPTV 374
```



```

;Title: Cloning, sequence, and photoregulation of al-1, a carotenoid biosynthetic gene
;Reference number: A35919; MUID:90377195; PMID:2144609
;Accession: A35919
;Status: preliminary
;Molecule type: DNA
;Residues: 1-595 <SCH>
;Cross-references: GB:M57465, GB:M33867; NID:G168746; PIDN:AAA33555.1; PID:Q168747
;Note: the authors translated the codon CTC for residue 157 as Glu, and GAG for residue 158 as Asp.
;Superfamily: phytoene dehydrogenase
;Keywords: transmembrane protein

```

[illegible]

RESULT 12

36968
 probable phytoene dehydrogenase (phytoene desaturase) - Streptomyces coelicolor
 ;Species: Streptomyces coelicolor
 ;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 ;Accession: T36968
 ;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1999

;Reference number: Z21607

;Accession: T36968
;Status: preliminary; translated from GR/EMBL/JDBET

Molecule type: DNA

;Residues: 1-523 <SEE>

;Cross-references: EMBL

```

;Experimental source: strain A3 (2)
;Corrections:

```

Gene: crte; SCQEDB:SCJ1.35

Superfamily: phytoene dehydrogenase

1000

Query Match	Best Local Similarity	Best Global Similarity	Score
24.2%	32.8%	32.8%	32.8%

UNITED STATES DEPARTMENT OF AGRICULTURE
BUREAU OF PLANT INDUSTRY
WASHINGTON, D. C.

Matches	161;	Conservative	76;	Mismatches	235;	Indels	19;	Gaps
QY	3	PT---VTVGAGFGGLAALAIPLQAAQATPVLLLEQRDKPGGRAYVYQGGTFDFAGATVITD	59					
Db	7	PTGHHVVVVGAGLSGLCALCAGRRVTVVRDAGFGGSGRVLGGYELDTGPTVLTLM	66					
QY	60	PSAIEELPALAGKOLKDYVELLPVTIPYRLCWESGVFNVDNDQALQEAQIQOF-NPRDV	118					
Db	67	PHLADAEFAAVGDSLRRRVELTALDAYRACPADGSGALDVHTDGEAMEAEVRRFAGPAQA	126					
QY	119	AGYRAFIDYSRAVFNQYKLGTVVFLSPFKMDLRAAPQLAKLQA---WRSVYSKVAGYIE	175					
Db	127	AGYRLRLRWLERLYRAQWRRFITDINFDSPLQLLH--PDLARLAALGGFGRLDGRIGRFLS	184					
QY	176	DEHLRQAFSPHSLVGCNRPATSSITYTLIALEREMGVWPPRGGTGALVNGMKILKFPDQLG	235					
Db	185	DERLRVFSQALLVAGVAPARALAAVAYATMDTVAGVWFFKGMGHALPRAMADAAATAG	244					
QY	236	GEVLNARVSHMETVGDKIQAQVLEDGRFPETCAVASNADVVHTTDRDLLSQHPAAAKQAK	295					
Db	245	ADLRASAEVKALERSAGRVRAVHLASGERIACDAVYLTCELSTAY-GLLGRAP-----R	297					
QY	296	KLQSKBMSNSLFLVYFGLNHHHDQLAHTTCVCFPRYRELIHEIFNHDGLAEPSLYLHAP	355					
Db	298	RPARLRHSFSAVILHAGTDITWPHLAHTLSFGAAWERTFELTRTGLAMSDPSLLITRP	357					
QY	356	CVTDPSLAPEGCGSYVYLAVPVPH--LGTANLDWAVEGPRLRDRIFDYLBQHWYMPGLRSOL	413					
Db	358	TTHPDPAAPPGRHLHYVLPCPNTDVGPDAAVARDLGPYRBSLVGLSRESGLGEPADSV	417					
QY	414	VTRHMTFTFDRDLNAWQCSASVPFILTQSAWPPPHNRDXHIINLXLVAGTAGTTPGAGI	473					
Db	418	QBELVTPLDWDAGGHA-AGSPFSVSHTPAQTGFFPRPNLVKGLNDVILVAGCGTTPGGVV	476					
QY	474	PGVIGSAKATA	484					

RESULT 13

B90061
squalene synthase [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 02-Aug-2002
C:Accession: B90061
R:Auroda, M.; Onta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, R.; Kuroda, M.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Seki, M.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A98758; PMID:21311952; PMID:11418146
A:Accession: B90061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-502 <RUP>
A:Cross-references: GB:BA000018; PTD:gl3702511; PTDN:BA043652.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: crtN
C:Superfamily: phytoene dehydrogenase

Query Match 23.9%: Score 624; DB 2; Length 502;

Best Local Similarity 31.7%; Pred. No. 4.4e-42;

Matches 159; Conservative 98; Mismatches 221; Indels

Q: 5 UTEBCECCI XEATRI CAACTIPUT.I.I.EORBDKRCGRAYVVOEOGETENAGPT

6 VIGAGFGGLTAAIRLQAAAGIFVLLLEQRDNFGGRAIVIQEQGTIFDAGE

Db 5 VIGAGVTGLAAAARIASQGHEVTIPEKNNNVGGRMNQLKKDGFTEDMGFTT

1000

Qy 66 LFALAGKQLKDYVELLEPTPFYRLCWESGKVFNYDNDQALEAQIQQFN

65 VFTACGKNVEDYIEELRQLYRYIVDYVFDHDDRTVFTDLAELOOMLESIEPP

SECRET

126 -----DYSRAVENEGVYKLGTVTFPLSEKMLRAAPQLAKIQAWRSVYKVGAYIEDSH 178
125 TDVYKVEIARIYFLERTYKPS-----DFYNM-TSVVQGAUKLTNLHADOLIEHYIDNEK 179
179 LRQAFSPHSLVGNPNPATSIYTLIHALREMGVWPPRGGTGALVNGMI:KLPQDLGGEV 238
180 IQKLLAFQTLVIGIDPKRPSLSYIIPMIEMFVGHFKGMYGMAQGLAQLAKNDLVNI 239
239 VLNARVSH--METVGDKIQAQVLE-DGRRFETCAVASNADVVHTYRDLLSOH-PAKAKQA 294
240 ELNAIEQIITDPKPKRADA:KLVNGDIRKFD--KILCTADPPSVASSLMDPFAPIKKYPP 297
295 KKLOSKRMSNSFLVLYFGLN-HHHDQLAHTVCFGRPYRELIIHFNHDLGLADFSLYLH 353
298 HKIADLDYSCAFLWYIGID:DVTDQVRLHNVIFSDDFRGNIIEIF-EGRLSYDPSIYV 356
354 APCVTDPSLAPEGCGSVYVLAIPVPHLT-ANLDWAVEG--PRLADRFIDYLEQ-HYMPGL 409
357 VPAVADKSLAPEGKGTGYVLMPTPELKTGSDWSDBALTQQIKHEIYKRLATIEVFEDI 416
410 RSQLVTHRMFTFPDFRDELNAWQSAFVSPEPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469
417 KSHVSTITFPNDFEQTYHAKFSA:GLAFTLAQSNYRPNQVSRDYKDYLPAGASTHP 476
470 GAGIPGVIGSKAKATAGLMLDL 491
477 GAGVPVILTSKATVDEMIKDI 498

SULT 14
5306
ytoene dehydrogenase (EC 1.3.-.-) - Myxococcus xanthus
Alternate names: phytoene desaturase
Species: Myxococcus xanthus
Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 29-Sep-1999
Accession: S35306; S27594
Fonfies, M.; Ruiz-Vazquez, R.; Murillo, F.J.
BO J. 12, 1265-1275, 1993
Title: Growth phase dependence of the activation of a bacterial gene for carotenoid sy
Reference number: S35306; MUID:93223667; PMID:8467787
Accession: S35306
Molecule type: DNA
Residues: 1-529 <FON>
Cross-references: EMBL:M94727; MID:g150079; PIDN:AAA25390.1; PID:g150080
Gene: carC
Superfamily: phytoene dehydrogenase
Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 23.68; Score 616; DB 2; Length 529;
Best Local Similarity 30.28; Pred. No. 2.1e-41;
Matches 153; Conservative 99; Mismatches 218; Indels 36; Gaps 9;
6 VIGAGFGGLALAIRLQAAGIPVLLLEQDKPKGRAYVYQGGFTF-----DAQPTVI 57
13 VVAGPGGLSAAINLACQGRVTVVERDAVPGGRW-----KGLTLCASGEAYVDTGPSIL 67
58 TDPSAIELFALAGKQLKDYVELLPVTFPRLCWESKGFVNDQAOLEAQIQGNPRD 117
68 QLPGVLEQIFRAARRLEDYVQLPLDVNTRVHFWDGTHLDTTRHLDREBAELAKFGRQ 127
118 VAGYRAFLYSRAVFNEGYKLL--GTPVFLSPKMLRAAPQLAKLQAWRSVYKVGAYIE 175
128 ASALRQWEDGREKYGIAYOKFICTSDNLGYAPWRLAFTL-RFPKWTLYRLQDGFPE 186
176 DEHURQAFSPHSLVGNPNPATSIYTLIHALREMGVWPPRGGTGALVNGMIKLPQDLG 235
187 DDRVYALAPSKYGLKHP:TTCSVSFVIFLEAFGVHVEGCFRELGRGMRCARDIG 246
236 GEVVLNARVSHMETVGGKIQAQVLE-EDGRRFETCAVASNADVVHTYRDLLSOH:PAKAKQA 295
247 ATFRMGTPVKRVADAGRVGVKLVGGEVLDAADVNNADLA:YARSLLI---PAEREGS 303

QY 296 K-----LOSKEVNSLFLVLYGLNHHHDQLAHTVCFGRPYRELIIHFNHDLG----- 344
DB 304 RUTDAALERAKYSCSTFMAYYGLDVTYADLPFHLLIYLSSESART-----DRDALEDRHVD 358
QY 345 AEDPSLYLHAFCVTDPSLAPEGCGSVYVLAIPVPHLGTANLDWAVEGFLRDRIDFYLEQH 404
DB 359 LEDPPYVNCVGTDPSCAPAGHSTLYLVPTNTGRP-VDMVKTQALRERI:PAKLEKV 417
QY 405 YMPGLRSOLVTHRMFTFPDFRDELNAWQSAFVSPEPILTQSAWFRPHNRDKHIDNLYLVG 464
DB 418 GLKGVREHIRERYFTATETWRDDFNVRGAVFNLSHTWLQGLPRLPKVKNORDIEGLYFVG 477
QY 465 AGTTPGAGIPGVIGSKAKATAGLMLDL 490
DB 478 GGTTPGSGELLTIMESANIADYLTRE 503

RESULT 15
AG2509
zeta-carotene desaturase [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2509
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-499 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA078339.1; PID:g17135793; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7255
A:Genome: plasmid
C:Superfamily: phytoene dehydrogenase

Query Match 23.28; Score 604; DB 2; Length 499;
Best Local Similarity 29.8%; Pred. No. 1.8e-40;
Matches 153; Conservative 89; Mismatches 224; Indels 48; Gaps 9;
QY 2 KPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPKGRAYVYQGGFTFDPAGTIVTDPS 61
DB 3 KKAIVAGAGPGGLATAIRLAGLGTQVEIFEAERVGRGMRGFEVDYAFDTGTPTILQLPH 62
QY 62 AIEELFALAGKQLKDYVELLPVTFPRLCWESKGFVNDQAOLEAQIQGNPR----- 116
DB 63 LYKELFEENLNPADYVQLKELAPYTELK?-----WDGTQDITSDLSQSFKTQATLR 115
QY 117 -DV-----AGYRAFL-DYSRAVFNEGYKLG-TVPFLSPKMLRAAPQL 157
DB 116 SDDLPAFDRWYSEHIRKYELGYKPYLAGPARSIF--GYLRPDELMEKFLSFR----- 164
QY 158 AKLQAWRSVYKVGAYIEDSHLRQAPSFHLLVGGNPFATSSYITLIHALREMGVWYFPR 217
165 -----PWNELTQHFRFPQDERLVYALSYPKYLGMFTVASSVFSIIPFLFSQGVNHPV 220
QY 218 GGTGALVNGMIKLPQDLGGEVVLNARVSHMETVGGKIQAQVLE-EDGRRFETCAVASNADVV 277
DB 221 GGFALAQGLANAQAQDLGVKIHLSHPVHQIWIIEQGVGLGLADASRHRQFDTVINADFA 280
QY 278 HTYRDLLSOH:PAKAKQAQKLOSKRMSNSFLVLYFGLNHHHDQLAHTVCGFPYRELIIHE 337
DB 281 YAVRHLLPTSARGYTDNKLQGMQFSCSTFMLYLGNNRRYEDLPHQIYILSDNIRLERP 340
QY 338 IFNHDLGLAE-DFSLYLHAPCVTDPSLAPEGCGSVYVLAIPVPHLGTANLDWAVEGFLRDR 396
DB 341 WYDSDALDETDPFYVNCNPTIIDPSNAPAGHS:TLFVLVPIPTSYA-VDMWDIKQKSYTDF 399
QY 397 IFDYLEOHYMPGLRSQLVTHRMFTFPDFRDELNAWQSAFVSPEPILTQSAWFRPHNRDKH 456

b 400 ILKRLHLLCHNIEQHIVTQSCYTAQSWLDDYRVHLCVAFNLSHNLTLQGPFRPIRSEN 459
Y 457 IDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLED 490
b 460 IAGLYWIGGAVHPGSGLLTLEASRSAGFIHQD 493

earch completed: February 29, 2004, 14:52:52
ob time : 15.6946 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:51:24 ; Search time 32.666 Seconds
(without alignments)
3180.293 Million cell updates/sec

tle: US-09-941-947A-32

irect score: 2606

quence: 1 MKPTTVIGAGFGGLAALRL

oring table: BLOSUM62

arched: Gapop 10.0 , Gapext 0.5

al number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*

17: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*

18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	2606	100.0	492	10	US-09-941-947A-32
2	2606	100.0	492	14	US-10-218-118-8
3	1802	89.1	491	15	US-10-369-493-234
4	1595	61.2	494	9	US-09-547-267-5
5	1595	61.2	494	10	US-09-920-923-4
6	1464	56.2	498	15	US-10-369-493-10579
7	1461	56.1	506	15	US-10-369-493-17830
8	1183.5	45.4	498	15	US-10-369-493-20438
9	1051	40.3	548	15	US-10-369-493-540
10	1014	38.9	498	15	US-10-369-493-19248
11	962.5	36.9	518	15	US-10-369-493-7750
12	793.5	30.4	511	15	US-10-369-493-19519
13	703.5	27.0	511	9	US-09-934-903-16
14	703.5	27.0	511	9	US-09-934-868-76
15	703.5	27.0	511	10	US-09-941-947A-22

Sequence 4, Appli
Sequence 18, Appl
Sequence 78, Appl
Sequence 24, Appl
Sequence 8, Appl
Sequence 8951, Ap
Sequence 12, Appl
Sequence 8562, Ap
Sequence 8142, Ap
Sequence 18983, A
Sequence 3769, Ap
Sequence 21541, A
Sequence 18644, A
Sequence 14, Appl
Sequence 18, Appl
Sequence 19547, A
Sequence 7895, Ap
Sequence 3, Appl
Sequence 18646, A
Sequence 10196, A
Sequence 5696, Ap
Sequence 4194, Ap
Sequence 19518, A
Sequence 20440, A
Sequence 19986, A
Sequence 5, Appl
Sequence 6, Appl
Sequence 2831, Ap
Sequence 18878, A
Sequence 17259, A

511 14 US-10-358-917-4
497 9 US-09-934-903-18
497 9 US-09-934-868-78
497 10 US-09-941-947A-24
497 14 US-10-358-917-8
494 15 US-10-369-493-8951
502 14 US-10-358-917-12
513 14 US-10-156-761-8562
506 15 US-10-369-493-8142
499 15 US-10-369-493-18983
604 23.6
602 23.1
514 15 US-10-369-493-3769
512 15 US-10-369-493-21541
512 15 US-10-369-493-18644
497 14 US-10-358-917-14
530 14 US-10-128-713A-18
404 15 US-10-369-493-19547
495 15 US-10-369-493-7895
1268 13 US-10-438-784-3
536 15 US-10-369-493-18646
498 15 US-10-369-493-10196
439 9 US-09-815-242-5696
548 9 US-09-738-626-4194
485 15 US-10-369-493-19518
496 15 US-10-369-493-20440
537 15 US-10-369-493-19986
469 15 US-10-438-784-5
542 14 US-10-209-372-6
542 15 US-10-369-493-2831
539 15 US-10-369-493-18878
498 15 US-10-369-493-17259

ALIGNMENTS

RESULT 1

US-09-941-947A-32
; Sequence 32, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Piatagglio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: C11903 US NA
; CURRENT APPLICATION NUMBER: US/09/941.947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-32

Query Match 100.0%; Score 2606; DB 10; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.2e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKPTTVIGAGFGGLAALRLQAAGI PVLLLEQRDKPGGRAYVYQEGFTFDAGPTVITDP 60
Db 1 MKPTTVIGAGFGGLAALRLQAAGI PVLLLEQRDKPGGRAYVYQEGFTFDAGPTVITDP 60
Qy 61 SAIEELFALAGKQKDYVELLPVTFPRLCNESGKVFNTDNDQAQLEAQIQFNPDRVAG 120

db 61 SAIEELFALAGKOLKDYVLLPVTPTFYRLCWESGKVFYNDQAOLEAQIQCFNPRDVAG 120
121 YRAFLDYSAVFNVEGYLKGTVFPLSKDMLRAAPOLAKLQAWRSVYKVGAGYIDEHLR 180
121 YRAFLDYSAVFNVEGYLKGTVFPLSKDMLRAAPOLAKLQAWRSVYKVGAGYIDEHLR 180
181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240
181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGERFETCAVSNADVVHTYRDLSSHAPAAKQAKKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGERFETCAVSNADVVHTYRDLSSHAPAAKQAKKLOSK 300
301 RMSNSLFLVYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLFLVYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492
ESULT 2
S-10-218-118-8
Sequence 8, Application US/10218118
Publication No. US20030148319A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
FILE REFERENCE: CUI876 US NA
CURRENT APPLICATION NUMBER: US/10/218,118
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: 60/312,646
PRIOR FILING DATE: 2001-08-15
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 492
TYPE: PRT
ORGANISM: Pantoea stewartii
US-10-218-118-8
Query Match 100.0%; Score 2506; DB 14; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.2e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
b 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
Y 61 SAIEELFALAGKOLKDYVLLPVTPTFYRLCWESGKVFYNDQAOLEAQIQCFNPRDVAG 120
b 61 SAIEELFALAGKOLKDYVLLPVTPTFYRLCWESGKVFYNDQAOLEAQIQCFNPRDVAG 120
Y 121 YRAFLDYSAVFNVEGYLKGTVFPLSKDMLRAAPOLAKLQAWRSVYKVGAGYIDEHLR 180
b 121 YRAFLDYSAVFNVEGYLKGTVFPLSKDMLRAAPOLAKLQAWRSVYKVGAGYIDEHLR 180
Y 181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240
b 181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240

db 181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240
Y 241 NARVSHMETVGDKIQAQVLEDGERFETCAVSNADVVHTYRDLSSHAPAAKQAKKLOSK 300
db 241 NARVSHMETVGDKIQAQVLEDGERFETCAVSNADVVHTYRDLSSHAPAAKQAKKLOSK 300
Y 301 RMSNSLFLVYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
db 301 RMSNSLFLVYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
Y 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
db 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDYLEQHYMPGLRSOLVTHRMFT 420
Y 421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
db 421 PFDFRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
Y 481 KATAGLMLEDLI 492
db 481 KATAGLMLEDLI 492
RESULT 3
US-10-369-493-234
; Sequence 234, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 234
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-234
Query Match 59.1%; Score 1802; DB 15; Length 491;
Best Local Similarity 67.4%; Pred. No. 4.5e-173;
Matches 331; Conservative 66; Mismatches 94; Indels 0; Gaps 0;
Y 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
db 1 MKPTTVIGAGFGGLALAIRLOAAGIPVLLLEORDKPGGRAYVYQGGTFDAGPTVITDP 60
Y 61 SAIEELFALAGKOLKDYVLLPVTPTFYRLCWESGKVFYNDQAOLEAQIQCFNPRDVAG 120
db 61 SAIEELFALAGKOLKDYVLLPVTPTFYRLCWESGKVFYNDQAOLEAQIQCFNPRDVAG 120
Y 121 YRAFLDYSAVFNVEGYLKGTVFPLSKDMLRAAPOLAKLQAWRSVYKVGAGYIDEHLR 180
db 121 YRAFLDYSAVFNVEGYLKGTVFPLSKDMLRAAPOLAKLQAWRSVYKVGAGYIDEHLR 180
Y 181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240
db 181 QAFSFLHLLVGGNPFATSSYITLIHALEREWGVPFRGGTGAALVNGMIKLFQDLGGEVVL 240
Y 241 NARVSHMETVGDKIQAQVLEDGERFETCAVSNADVVHTYRDLSSHAPAAKQAKKLOSK 300
db 241 NARVSHMETVGDKIQAQVLEDGERFETCAVSNADVVHTYRDLSSHAPAAKQAKKLOSK 300
Y 301 RMSNSLFLVYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360
db 301 RMSNSLFLVYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVTDP 360

3 301 KMSNSLFIYFGLNKLHICLAHTTICFCNRYKELLDDIFECSNIPGCFSLYLHSPSVTDP 360
/ 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
3 361 TLAPPGCASYYVLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
/ 421 PDFRDELNAMOGSAFSEVPILITQSAWFRPHNRDHIENLYVAGTHPGAGIPGVIGSA 480
3 421 PDFRDELNAMOGSAFSEVPILITQSAWFRPHNRDHIENLYVAGTHPGAGIPGVIGSA 480
/ 481 KATAGLMLEDL 491
3 481 KATAGLMLEDL 491

RESULT 4
1-09-547-267-5
Sequence 5, Application US/09547267
Patent No. US20020147371A1
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547.267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
1-09-547-267-5

Query Match 61.2%; Score 1595; DB 9; Length 494;
Best Local Similarity 61.9%; Pred. No. 3.9e-152;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;
1 MKPTTVICAGFGGLALAIRLOQAAGIPVLLLEQORDKPGGRAYVYQEQGFTPDAGPTVTIDP 60
1 MSSAIVICAGFGGLALAIRLOQAAGIATTTVEARDKPGGRAYVNDQGHVFDAGPTVTIDP 60
61 SAIELEFALAGKQLKDYVELLPVTTFYRLCWESKGFVNDQQAQLEAQIQQFNPRDVAG 120
61 DSLRELWALSQPMERDVTLLPVSPFFYRLTWADGRSFEYVNDDELLIRQVASFNPADVDG 120

Query Match 61.2%; Score 1595; DB 9; Length 494;
Best Local Similarity 61.9%; Pred. No. 3.9e-152;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;
1 MKPTTVICAGFGGLALAIRLOQAAGIPVLLLEQORDKPGGRAYVYQEQGFTPDAGPTVTIDP 60
1 MSSAIVICAGFGGLALAIRLOQAAGIATTTVEARDKPGGRAYVNDQGHVFDAGPTVTIDP 60
61 SAIELEFALAGKQLKDYVELLPVTTFYRLCWESKGFVNDQQAQLEAQIQQFNPRDVAG 120
61 DSLRELWALSQPMERDVTLLPVSPFFYRLTWADGRSFEYVNDDELLIRQVASFNPADVDG 120

QY 121 YRAFLDYGRAVNEGYKLGTVPFLSPKDMRLAAPOLAKLOAMRSVYKVGAGYIDEDHLR 180
DB 121 YRRFDYAEVYREGYKLGITPTFLKQOMLNAPALNRLOAYRSVHSMWARFIQDPHLR 180
QY 181 QAFSPHLLVGCNPFATSSIIYTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSPHLLVGCNPFATSSIIYTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDGRREFETCAVSNADVVHTYRDLLSQHPAAAKQAKKIQSK 300
DB 241 NARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTTRGRRTKAILNRQ 300
QY 301 RMSNSLFLVYFGLNKHHDLAHTTVCPCPRYRELTHTFNEHDGLAEDPSLYLHAPCVTDP 360
DB 301 RMSNSLFLVYFGLNKHHDLAHTTVCPCPRYRELTHTFNEHDGLAEDPSLYLHAPCVTDP 360
QY 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
DB 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPGLRSQLVTHRMFT 420
QY 421 PDFRDELNAMOGSAFSEVPILITQSAWFRPHNRDHIENLYVAGTHPGAGIPGVIGSA 480
DB 421 PDFRDELNAMOGSAFSEVPILITQSAWFRPHNRDHIENLYVAGTHPGAGIPGVIGSA 480
QY 481 KATAGLMLEDL 491
DB 481 KATAGLMLEDL 491
RESULT 5
US-09-920-923-4
Sequence 4, Application US/09920923
Publication No. US20030022273A1
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/09/920,923
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923-4
Query Match 61.2%; Score 1595; DB 10; Length 494;
Best Local Similarity 61.9%; Pred. No. 3.9e-152;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;
1 MKPTTVICAGFGGLALAIRLOQAAGIPVLLLEQORDKPGGRAYVYQEQGFTPDAGPTVTIDP 60
1 MSSAIVICAGFGGLALAIRLOQAAGIATTTVEARDKPGGRAYVNDQGHVFDAGPTVTIDP 60
61 SAIELEFALAGKQLKDYVELLPVTTFYRLCWESKGFVNDQQAQLEAQIQQFNPRDVAG 120
61 DSLRELWALSQPMERDVTLLPVSPFFYRLTWADGRSFEYVNDDELLIRQVASFNPADVDG 120
QY 121 YRAFLDYGRAVNEGYKLGTVPFLSPKDMRLAAPOLAKLOAMRSVYKVGAGYIDEDHLR 180
DB 121 YRRFDYAEVYREGYKLGITPTFLKQOMLNAPALNRLOAYRSVHSMWARFIQDPHLR 180
QY 181 QAFSPHLLVGCNPFATSSIIYTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSPHLLVGCNPFATSSIIYTLHALEREWGVMPRGSTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDGRREFETCAVSNADVVHTYRDLLSQHPAAAKQAKKIQSK 300
DB 241 NARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTTRGRRTKAILNRQ 300

301 RMSNSLVLYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVITDP 360
301 RWSNSLVLYFGLSKRENLAHNSVIFGPYKLVNEIFNGPELPDGFBNYLHSPCVITDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLWAVEGPRLRDRIFDYLEQHYMPGLRSOLVTHRMFT 420
361 SLAPEGNSTHYVLAPVPHLGSADVDEAEPGYAEIIFELERRAIPDLRKLTWSRIFS 420
421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
421 PADFTSTLSAHHGSAFSEVPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPGVGSA 480
481 KATAGLMLEDL 491
481 KATAGLMLEDL 491

RESULT 6
S-10-369-493-10579
Sequence 10579, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 10579
LENGTH: 498
TYPE: PRT
ORGANISM: Spingomonas aromaticivorans
S-10-369-493-10579

Query Match 56.2%; Score 1464; DB 15; Length 498;
Best Local Similarity 56.4%; Pred. No. 7e-139;
Matches 277; Conservative 78; Mismatches 136; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGRAYVYQGGTFDAGPTVITDP 60
7 MKRACVIGAGFGGLALAIRLQSGGVQTTLEARDKPGRAYFMEKDGTFDGGPTVITDP 66
61 SAIEELFALAGKQKDYVELLPVTPPYRLCWESGKVFYNDNDQAQLAEQIQOFPNPRDVAG 120
67 ACLEELWALKGRHMADVELMPVMPYRLNWDGTFNEDYSNDSDASLRABIAKLEPADVAG 126
121 YRAFLDYSRAVFNEGYLKGTVPFLSPKDMRAAPQALAKLQAWRSYISKVAGYIEDEHLR 180
127 YDFLAYSAGVFEGYVKGVPFLDFASWIKAAALARYQAWRSYISWVSFVNEKLR 186
181 QAFSPHSLVGGNPPATSIYTLIHALEREWGVPFRGGTGALVNGMIKLFODLGGVWL 240
187 EAFSPHLLVGGNPMFTSAIYALIHKLKRDGGVWMAKGTNLEIEGMYTHFRIGSEVRI 246
241 NARVSHMETVGDKIQAQVLEDGRRRPFCVANSADVVHTYRLDLSCHPAAAKQAKLQSK 300
247 GDPVTEILTGRTVGTVEFTRSGSGCFDAVASNADVMHSYRDLSDNSHAKRRAKSLGK 306
301 RMSNSLVLYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVITDP 360
307 RFSPSLVFVHFEGTGWPGIPHEMILFGPRYKGLLDDIYTHGVLPQDFSIVLHPTVITDP 366
361 SLAPEGCGSYVVLAPVPHLGTANLWAVEGPRLRDRIFDYLEQHYMPGLRSOLVTHRMFT 420
367 SVAPEGNSTFYALVPVANNMKLPVNMDEVGPILEKRIILDEVRRLIPDIHTRIVTKFHYA 426

421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
427 PSDFSTLSAHLGSAFSEPLLTQSAWFRPHNRDDAISNFYLVGAGTHPGAGIPGVGSA 486
481 KATAGLMLEDL 491
487 KATAGLMLEDL 497

RESULT 7
US-10-369-493-17830
Sequence 17830, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 17830
LENGTH: 506
TYPE: PRT
ORGANISM: Spingomonas
US-10-369-493-17830

Query Match 56.1%; Score 1461; DB 15; Length 506;
Best Local Similarity 56.8%; Pred. No. 1.4e-138;
Matches 279; Conservative 75; Mismatches 137; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGRAYVYQGGTFDAGPTVITDP 60
15 MRAVWVIGAGFGGLALAIRLQSGAVDTTVEARDKPGRAYFMEKDGTFDGGPTVITDP 74
61 SAIEELFALAGKQKDYVELLPVTPPYRLCWESGKVFYNDNDQAQLAEQIQOFPNPRDVAG 120
75 DAKEELWALSGHDSIDSDVTLPVTPPYRLSWPDGTFDYNDDAVLAGIAKLDSADIA 134
121 YRAFLDYSRAVFNEGYLKGTVPFLSPKDMRAAPQALAKLQAWRSYISKVAGYIEDEHLR 180
135 YRFLDYSASVVEEGYVKGAKAFLDPSKMLKAAAPALMQHQAWSYISVSGFVNEKLR 194
181 QAFSPHSLVGGNPPATSIYTLIHALEREWGVPFRGGTGALVNGMIKLFODLGGVWL 240
195 EALSPTLLVGGNPMATSIYALIHKLKRDGGVWMAEGGTNRLIAGMITHFERLGGKVRI 254
241 NARVSHMETVGDKIQAQVLEDGRRRPFCVANSADVVHTYRLDLSCHPAAAKQAKLQSK 300
255 GDPVDEILTGRTVGTVEFTRSGSGCFDAVASNADVMHSYRDLSDNSHAKRRAKSLGK 314
301 RMSNSLVLYFGLNHHDDQLAHTVCFGPYREL IHEIFNHDGLAEDFSLYLHAPCVITDP 360
315 RFSPSLVFVHFGLRGTFDPIPHMILFGPRYKGLLDDIYTNVGWVSEDFSLYLHESASDP 374
361 SLAPEGCGSYVVLAPVPHLGTANLWAVEGPRLRDRIFDYLEQHYMPGLRSOLVTHRMFT 420
375 SVAPEGNSTFYALVPVPHLGPADWSQVGVLEBRIAEVORRLIPDLAERIVTKFHYA 434
421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
435 PTDPRDLAAHSGSPLEPILTQSAWFRPHNRDAISNLYFVGAGTHPGAGIPGVGSA 494
481 KATAGLMLEDL 491
495 KATAELMLGEL 505

RESULT 8

3-10-369-493-20438
Sequence 20438, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 20438
LENGTH: 498
TYPE: PRT
ORGANISM: Rhodospseudomonas palustris
-10-369-493-20438

Query Match 45.4%; Score 1183.5; DB 15; Length 498;
Best Local Similarity 47.5%; Pred. No. 1.6e-110;
Matches 230; Conservative 84; Mismatches 169; Indels 1; Gaps 1;
6 VIGAGFGGLALAIRLQAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDPSAIBE 65
15 VIGAGFGGLAALVRLGAGRYVTVLEKLDKAGGRAYVHKQDGSFAGPTVITVAPLFEE 74
66 LFAIAGKQLKDYVLLPVTTPYRLCWESGKVFYNDQAOLEAQIQOFPNPDVAGYRAFL 125
75 LWKLCGRMSDDITLKPMSPFYRIRFDGTHFYSDDRDAVLQIAKFCDDVPAYDREW 134
126 DYSRAVENEGYKLGTVFFLSFKDMLRAAPOLAKIQAWSVSVKAGYIIEHLQAFSP 185
135 AASHEIFKVGEEQDGFHSHFTDMLKIAFAMKLSIRSVYGLVAHFKDPKLRQVFSF 194
186 HSLVGGNPFATSIYTLIHALLREBWGVPFGTGTALVNGMIKLPDLGGEVYVNLNARYS 245
195 HPLLIGNPFWSSSVCLITVLEKQWVHSAWGSGTALVGLVNLIEGQNTIRYNQDVR 254
246 HMETVGDKIQAQVLEDCGRPETCAVASNADVHTYRDLLSQHPAAKQAKKLSQKMSNS 305
255 QIVVENGACGVKLADGEVTKADIVWSNADSASTYRYLLPPETRKRWTDKIEKSYMS 314
306 LFLVYFGLNHHHDLAHTTVCQPRYRELTHIEIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
315 LFWYFGTKRYEDVKHTILLGPRYKELISDIFSRYVAEDFSLYLHAPCVTDPSLAPE 374
366 CGGYYVVLAPVPHLGTANLDWAVEGPELRDRIIFYLEQHYMPGLRSQVTHRMFTPFDR 425
375 GCUTFYVLSVPNL-LGDTOWHTKAETRYASIAKMLGATVLPDLENQIATSKITTPIDQ 433
426 DELNAWQSAFSEVPILOTQSAWFRPNRKHIDNLVVGAGTHPGAGIPGVIGSAKATAG 485
434 DRUSSFGAAGFLEPVLWQSAWFRPNQSEDEVKRLYLVGAGTHPGAGLPGVLSARVLA 493

RESULT 9

3-10-369-493-540
Sequence 540, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 540
LENGTH: 548
TYPE: PRT
ORGANISM: Deinococcus radiodurans
US-10-369-493-540

Query Match 40.3%; Score 1051; DB 15; Length 548;
Best Local Similarity 43.5%; Pred. No. 4.8e-97;
Matches 229; Conservative 72; Mismatches 185; Indels 40; Gaps 9;
2 KPTTVIGAGFGGLALAIRLQAGIPVLLLEQDKPGGRAYVYQ-EQGTTFDAGPTVITDP 60
17 KVALVIGAGFGGLALAIRLQAGIPVLLLEQDKPGGRAYVYQ-EQGTTFDAGPTVITDP 76
61 SAIEELFA-----LAGKQLK-----YVELLPVTPFYRLCWES 93
77 HFIEELPALERDRAGLDADPYPPVSLGSRVKEGSGSPHTSRVYTLVPLPFYRIVFHD 136
94 GKVFYNDQAOLEAQIQOFPNPDVAGYRAFLDYSAVENEGYKLGTVPSLSPKMLRA 153
137 GTYFDYDGPSTQRQIAELAPGDLAGYERPHADAELFRGFLLELGYTHFGDVPTMLRV 196
154 APOLAKIQAWSVSVKAGYIIEHLQAFSFFHSLVGGNPFATSIYTLIHALLREBWG 213
197 VEDLLKLDVATLTSFTSKYFQSDKLQVFSFETLLVGGNPLSVPAIYAMHFFVKTWGI 256
214 WFRGSGTALVNGMIKLPDLGGEVYVNLNARYSMTVGD-----KIQAVQLEDCGRPETC 268
257 HYAMGSGTALVNGMIKLPDLGGEVYVNLNARYSMTVGD-----KIQAVQLEDCGRPETC 316
269 AVASNADVHTYRDLLSQHPAAKQAKKLSQKMSN--SLFVLYFGLNHHHDLAHTTVC 325
317 LVASNGDWANTY--LKVRSARLVNSDLAVKAASEMSLLVYFGRGGDDLPKHHNI 374
326 CFGPRYRELTHIEIFNHDGLAEDFSLYLHAPCVTDPSLAPEGCGSYVYVLPVPHLGTANLD 385
375 LLGPRYRELTHIEIFNHDGLAEDFSLYLHAPCVTDPSLAPEGCGSYVYVLPVPHLGTANLD 433
386 WAVEGPELRDRIIFYLEQHYMPGLRSQVTHRMFTPFDRDELNAWQSAFSEVPILOTQ 444
434 WVEGPELRDRIIFYLEQHYMPGLRSQVTHRMFTPFDRDELNAWQSAFSEVPILOTQ 493
445 SANFRPNRKHIDNLVVGAGTHPGAGIPGVIGSAKATAGLMELED 490
494 SAFFPRPNRKHIDNLVVGAGTHPGAGIPGVIGSAKATAGLMELED 539

CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 18248
LENGTH: 498
TYPE: PRT
ORGANISM: Myxococcus xanthus
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(498)
OTHER INFORMATION: unsure at all xaa 1
3-10-369-493-19248

Query Match 38.9%; Score 1014; DB 15; Length 498;
Best Local Similarity 44.0%; Pred. No. 2.3e-93;
Matches 210; Conservative 80; Mismatches 185; Indels 2; Gaps 2;

6 VIGAGFGLAALAIRLQAAIGIPVLLLEORDKPGGRAYVYVQSGTFOAGPTVITDPSATEE 65
8 VIGSGFGLAALAIRLAAAGRWTVVLERMDSFGGRANAFQDGTTFDAGPTVITCPhillEs 67
66 LFLAGKQLKDVYELLPTVTFYRLCWESGKVFNYYNDQAOLEAQIOOFNRPDVAGYAPFL 125
68 LWALGGQRMADHVLPVAPLYRMEFPGDSTEDYHTDREAMRQSVRLSPRDEAGFLALC 127
126 DYSRAVFNEGYLKLTVPFLSPFDMRLAAPOLAKLOAWRSVSKVAGYLDEHLLQAQSF 185
128 ARVERMYEAGIGPMTVPDVLSPAPTPPALVRDEAFRSMFGVSKGHINDELQOALLSF 187
186 HSLIVGNPNPAT-SSIVTTLIHALEREWGVPFRGGTGCALVNGMIKUFQDLGGEVNLARV 244
188 HPLLIVGSPPTASAVYTSIQVPRRWGAFFPVGGTGCALVRGLLESIGSGEVYRGSEV 247
245 SHMETVCDKIQAVOLEDRRPETCAVASNADVHTYRDLLSQHPAAKQAKKLOSKRWSN 304
248 TETALEGRKATGVRLGGTGLAADVVVSNADAATWYRILVPGEVRKHWTDERINARVSM 307
305 SLFVLYFGLNHHHDQAHHTVTCFGRPYSELKEIFNEDGLAEDFSLYLHAPCVTDPSLAP 364
308 SAF-LWYFGTTRQVPEVAHHHTLLFGKDFRGMTSGLEGPGQPSADPLLILHRPTATDAALP 367
365 ECGSYVYLAPVPHLGTANLDHVAVGPRLRDRIFDYLEQHYQMGLRSQLVHTRMTTPDF 424
368 SGHDAFVYLAAPVHLGTF-GAEWKQRAEAFRELEEARSRVTLVPLGSELMTRSMATPEVF 426
425 RDELNAWQSAFVSEILFQSAWFFPNRDKHIDNMLVIGAGTHPGAGIPGVIGSAK 481
427 PGRIRPEKGAFAFSAFTWOTTEFLQAQASRDVNCIVWVGCTHPCAGIPAVICSAK 483

RESULT 11
 3-10-369-493-7750
 sequence 7750, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 7750
 LENGTH: 518
 TYPE: PRT
 ORGANISM: Rhodobacter rubroaeroides

US-10-369-493-7750

Query Match	36.9%	Score	962.5	DB 15	Length	518			
Best Local Similarity	41.2%	Pred.	No. 3.9e-88						
Matches	200	Conservative	78	Mismatches	201	Indels	7	Gaps	5
QY	6	VIGAGPGGLAIRLOAAGIPVLLLRQDKPGRAVYVOEGGFTFDAGPTVITDPAIAIE	65						
DB	15	VIGSGUGLAAARIGAKWRVTVIDKLDPGGRGSSIQEGHRLFGLPTIVTPQSLRD	74						
QY	66	LFALAGKQLKDYVELLPVTPPYRLCLWESKVFYNDQALEAQIQFNPRDVAGYRAPL	125						
DB	75	LWKTGREDPADVELAKPIDPFYEVWRPDSGSHFTVRQSTEAMKAELVARLSBGDVAGYEKFL	134						
QY	126	DYSRAVFNNGYKLGTVVPFLSPFKMDLRAAPQAKLQAWRSVYISKVAGYIDEHLROAFSP	185						
DB	135	KDSEKRYWFGYEDLGRSRMGLWDLIKVLPTFGWMDRDSRYVHAALRVKDELRMLSLF	194						
QY	186	HSLLVCGNPATSSITYTLIHALERENGWVFPFGGTGALVNGMIKLPDQLGGEVVLNARVS	245						
DB	195	HPLFIGDPFNVTSMYLVSQLKEKRGVHAIGVAAIAAAMAKV7IEGGGSPRMTEYD	254						
QY	246	HMETVGDKIQAOLEDGRFRFETCAVASNADVVHTYRDLLSQHPAAAKQAKKLSKMSNS	305						
DB	255	EILVEKGTATGVELASGEVLKAGLVVSNADAGTYMRLLENHRRRTWDAHVKSRSWMSG	314						
QY	306	LFVLYFGLANHH--DOLAHHTVCPGPRYRELIHEIFNHDGLAEDSLYLHACVTDPSL	362						
DB	315	LFVWYFGTKGMPDVGHTTIVNAPRYKGLVEDFLKGLAKDMSLYTHRPSITDPTV	374						
QY	363	APGCCGSYYVLAPVPHLGTAN-LDMAVEGPRLEDRFDVLEQHYMPGLRSQLVTHRMFTF	421						
DB	375	APGDDTFFALSPVPHLKQAPVDMQVAEPYRESVLEVLQSG-MFGIGERIGPSLVFTF	433						
QY	422	FDPRDS-LNAWQGSAPFSVPPILTQSAWFRPHNRDKHIDNLYLVGAGTHPCAGIPGVIGSA	480						
DB	434	ETEDRYLSPW-GAGFSIEPRILQSAWFRPHNI SEEVANILFLVGAGTHPCGAPGVGVIGSA	492						
QY	481	KATAGL	486						
DB	493	EVNAKL	498						

```

RESULT 12
US-10-369-493-19519
? Sequence 19519, Application US/10369493
? Publication NO. US20030233675A1
? GENERAL INFORMATION:
? APPLICANT: Cao, Yongwei
? APPLICANT: Hinkle, Gregory J.
? APPLICANT: Slater, Steven C.
? APPLICANT: Goldman, Barry S.
? APPLICANT: Chen, Xianfeng
? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS
? TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
? FILE REFERENCE: 38-10520521B
? CURRENT APPLICATION NUMBER: US/10/369,493
? CURRENT FILING DATE: 2003-02-28
? PRIOR APPLICATION NUMBER: 039
? PRIOR FILING DATE: 2002-02-21
? NUMBER OF SEQ ID NOS: 47374
? SEQ ID NO 19519
? LENGTH: 517
? TYPE: PRT
? ORGANISM: Myxococcus xanthus
US-10-369-493-19519

```

Query Match 30.4%; Score 793.5; DB 15; Length 517;
Best local Similarity 36.1%; Pred. NO. 4.8e-71;
Matches 175; Conservative 89; Mismatches 214; Indels 7; Gaps 4;
Qy 6 VIGAGFGGLALIRLQAAGIPVLLLEQRDKPGGRATVYQGGTFPAGGTVIDPSAIRE 65

12 VVGAGVGGGLAAARLAHAGGFDVQVFEKQGGGRCNRLQVDFGTWIDGPTIVLMPEVFE 71
66 LALAGKQKDYVELLPVTTPYRLCWESGKVFYNDNDQALQIAQIQFNPDRVAGYRAFL 125
72 TFAVGRVIEDYTLRCDDPNRVHFRDGSVDTFTSBLCAMGRELVERPGSYARYLAFL 131
126 DYSAVFNHGYLKGTVFSLFSEKML--RAAPQAKLQAMRSVTSYKVGAYIEDEHLQAF 183
132 AQGRVQVFTSLDHLVGRNYAGRLDYLSPVLAR:FOVRAHRRMYADVSRFFODERLSAAM 191
184 SPHSLVGNPNPATSSIVTLIHALERENGVWFFRGGTGALVNGMKLFDQLGGEWVLNAR 243
192 TFOYMLGVSPYASPAVYGLPPTFELGVIPFPKGLYAIPOALERLARERGVFHYGAP 251
244 VSHMETVGDKIQAOLEGRFETCAVASNADVVHTYRDLISQHPAAKAKKLQSRMS 303
252 VERILTDGRTGRVLEGSEVYADAVLCNADLPYAYEKLDPRAATLTKRKEKL---RYT 308
304 NSLVLYFGLNHHDDQLAHTVCGPRYRELHRIENHDLGAEFSLYLHAPCVTDPSLA 363
309 SSGYMLYLGKRRYPELLHNHVFRDYKGSFDDIPEFRVPEPDPSPYVNAFTDASLA 368
364 PEGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIDYLEQHYMFLGSLVTHRMFTFFD 423
369 PEGKDALVYLVVEPH-QHPDLWDKVBGPKVRAKPFARMAELGFPSPLESIDIEVERVFTPD 427
424 FDEELNAGQSASFVEPILTQSANFRPHNRDKHDNLVYLCAGTHPGAGIPGVIGSAK-A 482
428 WAGTFNLARGSAFGLSQNFQTQIGFPRPSNQARVKNLFFVGASTQPGTGLPTVLISARLV 487
483 TAGLM 487
488 TERLM 492

SULT 13
US-09-934-903-16
Sequence 16, Application US/09934903
Patent No. US20020102690A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odom, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. US20020102690A1, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: CL1646 US NA
CURRENT APPLICATION NUMBER: US/09/934,903
CURRENT FILING DATE: 2001-08-22
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 16
LENGTH: 511
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF8
US-09-934-903-16

Query Match 27.0%; Score 703.5; DB 9; Length 511;
Best Local Similarity 31.7%; Pred. No. 6e-62;
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;
2 KPTTVIGAGFGLALAIQLAAGIPVLLLEQDKPGGRAYVYQGGTFDAGPTVITDPS 61
5 KHIIIVAGPGLCAGMLLSQSGFKVSIFDKHAIIGGNRPINNNGFTFDGPTFLMKG 64
62 ATEELFALAGKQKDYVELLPVTTPYRLCWESGKVFYNDNDQALQIAQIQFNPDRVAGY 121

65 VLDEMFEICERSESDYLEFLPSMYRLLYDDRDIFVY--SDRENRAELQVFEDEGTDGY 123
122 RAFLDYSRAVFNHGY-----LKGTVFPLSPKMLRAAPQAKLQAMRSVTSYKVG 172
124 EQFMEQERKRFNALYPCITRDYSSLK-----SFLSL-DLIKALPWL---PFSVFNMLGQ 175
173 YIEDHSLQAFSFLHLYGVGNPFATSSIVTLIHALERENGVWFFRGGTGALVNGMKLQ 232
176 YFNOEKMLAFQFQSKYLGMSFWPCPALFTMLPYLEHEYGYIYHVKGGLNR:IAAAMAQVIA 235
233 DLGGSEVLNARVSHMETVGDKIQAOLEGRFETCAVASNADVVHTYRDLISQHPAAK 292
236 ENGGEIHLNHSIESLIIENGAAGVKVLOHGAELGDEVIINADFAHAMTHLVK--PGVLK 293
293 Q--AKLQSKMSNSLFLVYFGLNHHDDQLAHTVCGPRYRELHRIENHDLGAEFSL 350
294 KYTPENLKQREYSYSTFMLYGLGDKIYD-LPHHTIVFAKDYTTNIRNIFDKTKLTDPSF 352
351 YLHAFCTVDPSLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIDYLEQHY-MEGL 409
353 YVONASASDDSLAPAKGSAALYVLVPMN--NDSGLDWAQHQCQVREQVLDLIGALGLSDI 411
410 RSQLVTHRMFTFFDPRDSLNAGQSASFVEPILTQSANFRPHNRDKHDNLVYLCAGTHP 469
412 RAHIECEKIIITPQWETDEHYVKGATFSLSHKPSQMLYWRPHNRPHNFEELANCYLVGGGTHP 471
470 GAGIPGVIGSAKATAGLM 487
472 GSGLEPTIYESARISAKLI 489

RESULT 14
US-09-934-868-76
Sequence 76, Application US/09934868
Patent No. US20020137190A1
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odom, James M
APPLICANT: Schenzle, Andreas J
TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
FILE REFERENCE: CL1596 US NA
CURRENT APPLICATION NUMBER: US/09/934,868
CURRENT FILING DATE: 2001-08-22
PRIOR FILING DATE: 60/229,858
NUMBER OF SEQ ID NOS: 81
SOFTWARE: Microsoft Office 97
SEQ ID NO 76
LENGTH: 511
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by CRTN1
US-09-934-868-76

Query Match 27.0%; Score 703.5; DB 9; Length 511;
Best Local Similarity 31.7%; Pred. No. 6e-62;
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;
2 KPTTVIGAGFGLALAIQLAAGIPVLLLEQDKPGGRAYVYQGGTFDAGPTVITDPS 61
5 KHIIIVAGPGLCAGMLLSQSGFKVSIFDKHAIIGGNRPINNNGFTFDGPTFLMKG 64
62 ATEELFALAGKQKDYVELLPVTTPYRLCWESGKVFYNDNDQALQIAQIQFNPDRVAGY 121
65 VLDEMFEICERSESDYLEFLPSMYRLLYDDRDIFVY--SDRENRAELQVFEDEGTDGY 123
122 RAFLDYSRAVFNHGY-----LKGTVFPLSPKMLRAAPQAKLQAMRSVTSYKVG 172
124 EQFMEQERKRFNALYPCITRDYSSLK-----SFLSL-DLIKALPWL---PFSVFNMLGQ 175
173 YIEDHSLQAFSFLHLYGVGNPFATSSIVTLIHALERENGVWFFRGGTGALVNGMKLQ 232

176 YFNQKRLAPCFQSKYLGMSWSPALFTMLPYLHEHYGYHVKGLNRIIAAQAQVIA 235
 233 DLGGEVLNARVSHMETVGDKIQAQVLEDDGRRFETCAVASNADVVHTYRDLLSQHFAAAK 292
 236 ENGGEHLNSELIESLIIENGAAGKVKLQGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
 293 Q--AKKLOSKRMSNSLFVLYFGLNHHHDLQAHHTVCGPRYRELHETFNHDDGLAEDFSL 350
 294 KYPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTTNIRNIFDNKLTDDFSP 352
 351 YLHAPCVTPDPSLAPEGCGSYVYVLAHPHGLTANLDWAVEGPRLEDRIDFYLEQHY-MPGL 409
 353 YQNASASDLSAPAGKSALYVLPMPN-NDGLDQAHQCNVREQVLTLCARGLSDI 411
 410 RSLVTHRMFTFPDFDELNAWQGSAPSVPEPILTOSAMFRPHNRKHDIDNLYVCGAGTHP 469
 412 RAHIECKEITPTQWETDEHVYKGFATFSLSHKFSQMLYWRPHNRPFEELANCYLVGCGTHP 471
 470 GAGIPGVIGSAKATAGLM 487
 472 GSGLPTIYESARISAKLI 489

RESULT 15
 S-09-941-947A-22
 Sequence 22, Application US/09941947A
 Publication No. US20030003528A1
 GENERAL INFORMATION:
 APPLICANT: Brzostowicz, Patricia C.
 APPLICANT: Cheng, Qiong
 APPLICANT: DiCosimo, Deana J.
 APPLICANT: Koffas, Mattheos
 APPLICANT: Miller, Edward S. Jr.
 APPLICANT: Odom, J. Martin
 APPLICANT: Picataggio, Steve
 APPLICANT: Rouviere, Pierre E.
 TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
 FILE REFERENCE: CL1903 US NA
 CURRENT APPLICATION NUMBER: US/09/941,947A
 CURRENT FILING DATE: 2001-09-01
 PRIOR APPLICATION NUMBER: 60/229,907
 PRIOR FILING DATE: 2000-09-01
 PRIOR APPLICATION NUMBER: 60/229,858
 PRIOR FILING DATE: 2000-09-01
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 22
 LENGTH: 511
 TYPE: PRT
 ORGANISM: Methylobionas 16a
 IS-09-941-947A-22
 Query Match 27.0%; Score 703.5; DB 10; Length 511;
 Best Local Similarity 31.7%; Pred. No. 6e-62;
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;
 2 KPTTVIGAGGGALAIRQAAGIPVLLLEQORDKPGRAVYVQEGFTEDAGTVIDTPS 61
 5 KHIIIVAGGPGGLCAGKLLSQRGFKVSIFDKHAEIGGRNRPINMNGFTPTGTFLMKG 64
 62 AIELFALAGKQKDYVELLPVTPFYRLCWESGKVFNYDNDQAQLEAQIQQFNPRDVAGY 121
 65 VLDEMFELCERSEDIYLFPLSPMYRLLYDDRDIFVY-SDRENWGAELQRVFDEGTGY 123
 122 RAFLDYGRAVFNQY-----LKIGTVPLSFYKMDLRAAPOLAKQAWRSVYKVG 172
 124 EQFMEQRKRFNALYPCITRDYSSLK----SFLSL-DLIKALPWL-PPKSVFNLLQ 175
 173 YIEDEHLRQAFSFLVGVGNPPATSIYTLIALEREMGWPPRGCTGALVNGMIKLFQ 232
 176 YFNQKRLAPCFQSKYLGMSWSPALFTMLPYLHEHYGYHVKGLNRIIAAQAQVIA 235

233 DLGGEVLNARVSHMETVGDKIQAQVLEDDGRRFETCAVASNADVVHTYRDLLSQHFAAAK 292
 236 ENGGEHLNSELIESLIIENGAAGKVKLQGAELRGDEVIINADFAHAMTHLVK--PGVLK 293
 293 Q--AKKLOSKRMSNSLFVLYFGLNHHHDLQAHHTVCGPRYRELHETFNHDDGLAEDFSL 350
 294 KYPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTTNIRNIFDNKLTDDFSP 352
 351 YLHAPCVTPDPSLAPEGCGSYVYVLAHPHGLTANLDWAVEGPRLEDRIDFYLEQHY-MPGL 409
 353 YQNASASDLSAPAGKSALYVLPMPN-NDGLDQAHQCNVREQVLTLCARGLSDI 411
 410 RSLVTHRMFTFPDFDELNAWQGSAPSVPEPILTOSAMFRPHNRKHDIDNLYVCGAGTHP 469
 412 RAHIECKEITPTQWETDEHVYKGFATFSLSHKFSQMLYWRPHNRPFEELANCYLVGCGTHP 471
 470 GAGIPGVIGSAKATAGLM 487
 472 GSGLPTIYESARISAKLI 489

Search completed: February 29, 2004, 15:28:02
 Job time : 34.666 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 14.951 Seconds
(without alignments)
1698.885 Million cell updates/sec

US-09-941-947a-32

fect score: 2506

quence: 1 MKPTTVICAGFGGLAIAIRL.....IPGVIGSAKATGLMLDLII 492

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Issued Patents AA.*
1: /cgm2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgm2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgm2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgm2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgm2_6/ptodata/2/iaa/6CTUS COMB.pcp.*
6: /cgm2_6/ptodata/2/iaa/6CTUS COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2438	93.6	492	1	US-07-783-705A-4
2	1849.5	71.0	489	1	US-08-095-726-8
3	1849.5	71.0	489	1	US-08-096-043-8
4	1849.5	71.0	489	1	US-08-096-623A-8
5	1845.5	70.8	489	1	US-08-095-726-10
6	1845.5	70.8	489	1	US-08-096-043-10
7	1845.5	70.8	489	1	US-08-096-623A-10
8	1595	61.2	494	3	US-08-660-645A-5
9	1595	61.2	494	3	US-09-298-718-5
10	1595	61.2	494	3	US-09-548-969-5
11	1595	61.2	494	3	US-08-980-832-4
12	1595	61.2	494	4	US-09-547-267-5
13	1595	61.2	494	4	US-09-920-923B-4
14	703.5	27.0	511	4	US-09-934-903-16
15	687	26.4	497	4	US-09-934-903-18
16	582	21.6	582	4	US-09-091-725-17
17	259.5	10.0	290	4	US-08-936-165A-462
18	201	7.7	610	4	US-09-443-184-55
19	140	5.4	527	1	US-08-132-168A-32
20	134	5.1	544	1	US-08-472-028A-8
21	134	5.1	544	2	US-08-808-931-8
22	134	5.1	544	3	US-08-808-323-8
23	134	5.1	544	3	US-09-050-603A-8
24	134	5.1	544	3	US-09-102-420B-8
25	134	5.1	544	3	US-09-071-296-8
26	134	5.1	544	3	US-09-196-268-8
27	134	5.1	544	3	US-09-015-683-8

28	134	5.1	544	4	US-09-191-998-8
29	134	5.1	544	4	US-09-437-928-8
30	131	5.0	458	4	US-09-328-352-7549
31	130	5.0	414	3	US-09-067-626-4
32	121.5	4.7	571	4	US-09-690-942-15
33	118	4.5	409	4	US-09-252-991A-24944
34	118	4.5	474	4	US-09-543-681A-6873
35	118	4.5	614	4	US-09-567-003C-22
36	116.5	4.5	524	3	US-09-210-388-1
37	116.5	4.5	524	3	US-09-912-176-1
38	115.5	4.4	578	4	US-09-690-942-10
39	115	4.4	508	1	US-08-472-028A-4
40	115	4.4	508	2	US-08-808-931-4
41	115	4.4	508	3	US-08-808-323-4
42	115	4.4	508	3	US-09-050-603A-4
43	115	4.4	508	3	US-09-102-420B-4
44	115	4.4	508	3	US-09-071-296-4
45	115	4.4	508	3	US-09-196-268-4

ALIGNMENTS

RESULT 1
US-07-783-705A-4
; Sequence 4, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, Shigeyuki
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOID
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: N/A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/783,705A
FILING DATE: 19911023
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-103078
FILING DATE: 21-APR-1989
APPLICATION NUMBER: JP 2-53225
FILING DATE: 05-MAR-1990
APPLICATION NUMBER: US 07/519,011
FILING DATE: 13-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schwadron, Janet I.
REGISTRATION NUMBER: 33,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-708-1935
TELEFAX: 212-246-5959
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-705A-4

Query Match 93.6%; Score 2438; DB 1; Length 492;

Best Local Similarity 93.5%; Pred. No. 1e-234;
Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDKGGRAYVYQOGFTFDAGPTVITDP 60
1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDKGGRAYVYQOGFTFDAGPTVITDP 60

61 SAIEELPALAGKOLKDYVELLPVTPPYRLCWESGKVFYNDQOAEQIQFNPDRVAG 120
61 SAIEELPALAGKOLKDYVELLPVTPPYRLCWESGKVFYNDQOAEQIQFNPDRVAG 120

121 YRAFLDYSRVAFNEGKLVGTPFLSPKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
121 YRAFLDYSRVAFNEGKLVGTPFLSPKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180

181 QAFSFLSLVGGNPPATSSITTYLIHALEREWGVPFRGGTGALVNGMIKLFODLGGEVVL 240
181 QAFSFLSLVGGNPPATSSITTYLIHALEREWGVPFRGGTGALVNGMIKLFODLGGEVVL 240

241 NARVSHMETVGDKIQAQVLEDGRFRFTCAVASNADVVHTYRDLLOHPAAAKQAKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGRFRFTCAVASNADVVHTYRDLLOHPAAAKQAKLOSK 300

301 RVNSLFLVYFGLNHHDDQLAHETVCFGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360
301 RVNSLFLVYFGLNHHDDQLAHETVCFGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360

361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLECHYMPGLRSOLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLECHYMPGLRSOLVTHRMFT 420

421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

481 KATAGLMEDLI 492
481 KATAGLMEDLI 492

RESULT 2
US-08-095-726-8
Sequence 8, Application US/08095726
Patent No. 5530188
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Huei-Che B
TITLE OF INVENTION: Beta-Carotene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
PRIORITY APPLICATION DATA:
PRIORITY APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530188val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-726-8

Query Match 71.0%; Score 1849.5; DB 1; Length 489;
Best Local Similarity 72.3%; Pred. No. 6.1e-176;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

QY 1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDKGGRAYVYQOGFTFDAGPTVITDP 60
DB 1 MKPTTIVGAGGGLAALRLQAAGIPVLLLEQKDKGGRAYVYQOGFTFDAGPTVITDP 60

QY 61 SAIEELPALAGKOLKDYVELLPVTPPYRLCWESGKVFYNDQOAEQIQFNPDRVAG 120
DB 61 SAIEELPALAGKOLKDYVELLPVTPPYRLCWESGKVFYNDQOAEQIQFNPDRVAG 120

QY 121 YRAFLDYSRVAFNEGKLVGTPFLSPKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180
DB 121 YRAFLDYSRVAFNEGKLVGTPFLSPKDMRAAPOLAKLQAWRSVYKVGAYIEDEHLR 180

QY 181 QAFSFLSLVGGNPPATSSITTYLIHALEREWGVPFRGGTGALVNGMIKLFODLGGEVVL 240
DB 181 QAFSFLSLVGGNPPATSSITTYLIHALEREWGVPFRGGTGALVNGMIKLFODLGGEVVL 240

QY 241 NARVSHMETVGDKIQAQVLEDGRFRFTCAVASNADVVHTYRDLLOHPAAAKQAKLOSK 300
DB 241 NARVSHMETVGDKIQAQVLEDGRFRFTCAVASNADVVHTYRDLLOHPAAAKQAKLOSK 300

QY 301 RVNSLFLVYFGLNHHDDQLAHETVCFGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360
DB 301 RVNSLFLVYFGLNHHDDQLAHETVCFGPRYRELIEIFNHDGLAEDFSLYLHAPCVITDP 360

QY 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLECHYMPGLRSOLVTHRMFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRLEDRIFDYLECHYMPGLRSOLVTHRMFT 420

QY 421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDLNAWQGSASFVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

QY 481 KATAGLMEDLI 492
DB 481 KATAGLMEDLI 492

RESULT 3
US-08-096-043-8
Sequence 8, Application US/08096043
Patent No. 5530189
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Huei-Che B
TITLE OF INVENTION: Lycopene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL

COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096.043
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,568
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
:-08-096-043-8

Query Match 71.0%; Score 1849.5; DB 1; Length 489;
Best Local Similarity 72.3%; Pred. No. 6.1e-176;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLALALRLOAGIPVLLLSQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALALRLOAGIPVLLLSQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
61 SAIEELPALAGKQKDYVELLPVTPFYLCHESKVFYNDQALQAEQIQFPRVAG 120
61 TALEALFTLAGRMEDYVRLLPVKPFYRLCHESKVFYNDQALQAEQIQFPRVAG 120
121 YRAFLDYGRAVNEGYLGTVPFLSKDMLRAAPOLAKLQAEVSVSKVAGYIEDEHLR 180
121 YRFLAYSCAVFQEGYRLGSPVFLSFRDMLRAGFOLLKLAQMSVYQSVSEFIEDEHLR 180
181 QAFSPHSLVGVGNPFATSSIVYLIALEREWGVPPRGCTGALVNGMIKLPQDLGGEVVL 240

350LT 4
3-08-096-623A-8
Sequence 8, Application US/08096623A
Patent No. 5684238
GENERAL INFORMATION:

1 MKPTTVIGAGFGGLALALRLOAGIPVLLLSQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALALRLOAGIPVLLLSQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
61 SAIEELPALAGKQKDYVELLPVTPFYLCHESKVFYNDQALQAEQIQFPRVAG 120
61 TALEALFTLAGRMEDYVRLLPVKPFYRLCHESKVFYNDQALQAEQIQFPRVAG 120
121 YRAFLDYGRAVNEGYLGTVPFLSKDMLRAAPOLAKLQAEVSVSKVAGYIEDEHLR 180
121 YRFLAYSCAVFQEGYRLGSPVFLSFRDMLRAGFOLLKLAQMSVYQSVSEFIEDEHLR 180
181 QAFSPHSLVGVGNPFATSSIVYLIALEREWGVPPRGCTGALVNGMIKLPQDLGGEVVL 240

Query Match 71.0%; Score 1849.5; DB 1; Length 489;
Best Local Similarity 72.3%; Pred. No. 6.1e-176;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLALALRLOAGIPVLLLSQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALALRLOAGIPVLLLSQRDKPGRAVYVQEQGFTFDAGPTVITDP 60
61 SAIEELPALAGKQKDYVELLPVTPFYLCHESKVFYNDQALQAEQIQFPRVAG 120
61 TALEALFTLAGRMEDYVRLLPVKPFYRLCHESKVFYNDQALQAEQIQFPRVAG 120
121 YRAFLDYGRAVNEGYLGTVPFLSKDMLRAAPOLAKLQAEVSVSKVAGYIEDEHLR 180
121 YRFLAYSCAVFQEGYRLGSPVFLSFRDMLRAGFOLLKLAQMSVYQSVSEFIEDEHLR 180
181 QAFSPHSLVGVGNPFATSSIVYLIALEREWGVPPRGCTGALVNGMIKLPQDLGGEVVL 240

181 QAFSHLLVGGNPF... 240
241 NARVSHMETGDKI... 300
241 NARVEELVADNRV... 300
301 RMSNSLFVYFGLNHH... 360
301 SMSNSLFVYFGLN... 360
361 SLAPEGCGSVYVLA... 420
361 SLAPPCASFVYLA... 420
421 PFDFRDELNAWQ... 480
421 RQTSRHAWIALG... 477
481 KATAGLMEIDL 491
478 ESTASLMEIDL 488

RESULT 5
US-08-095-726-10
Sequence 10, Application US/08095726
Patent No. 5530188
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Profitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Beta-Carotene Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530188val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-726-10
Query Match 70.8%; Score 1845.5; DB 1; Length 489;
Best Local Similarity 72.1%; Pred. No. 1.5e-175;
Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

QY 1 MKPTVTIGAGGGLA... 60
DB 1 MEKTVVIGAGGGLA... 60
QY 61 SAIBELPALAGKOL... 120
DB 61 TALEALFTLAGRME... 120
QY 121 YRAFLDYSRAVEN... 180
DB 121 YRRFLAYSQAVF... 180
QY 181 QAFSHLLVGGNPF... 240
DB 181 QAFSHLLVGGNPF... 240
QY 241 NARVSHMETGDKI... 300
DB 241 NARVEELVADNR... 300
QY 301 RMSNSLFVYFGL... 360
DB 301 SMSNSLFVYFGL... 360
QY 361 SLAPEGCGSVYVLA... 420
DB 361 SLAPPCASFVYLA... 420
QY 421 PFDFRDELNAWQ... 480
DB 421 RQTSRHAWIALG... 477
QY 481 KATAGLMEIDL 491
DB 478 ESTASLMEIDL 488
RESULT 6
US-08-096-043-10
Sequence 10, Application US/08096043
Patent No. 5530189
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Profitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Lycopen Biosynthesis in
TITLE OF INVENTION: Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,043
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,568
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B

TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
1-08-096-043-10

Query Match 70.8%; Score 1845.5; DB 1; Length 489;
Best Local Similarity 72.1%; Pred. No. 1.5e-175;
Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDP 60
61 SAIEELFALAGKQKDYVELLPVTFYELCWESGKVFYNDQALQAEQAOIQFNPRDVAG 120
61 TALEALFTLAGRMEDYVRLLPVTFYELCWESGKTYLDYANDSFELEAQITQFNPRDVAG 120
121 YRAFLDYGRAVFNEGYLKIGTVPFSLFKDMLRAAPQLAKQAWRSVYSKVAGYIEDEHLR 180
121 YRRFLAYSQAVFQEGYLRGLSVFSLFRDMLRAGPQLLKQAWQSVYQSVSRFIEDEHLR 180
181 QAFSFSLLVGNPPTATSSITLIHALRENGVWPPRGCTGALVNGMKLPDGLGGEVVL 240
181 QAFSFSLLVGNPPTATSSITLIHALRENGVWPPRGCTGALVNGMKLPDGLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDDGRRETCVASNADVHTYRDLLSQHPAAKQAKLQSK 300
241 NARVVELVADNRVSVQRLADGRIFEDTAVASNADVHTYKLLGTIPVQKRAARLERK 300
301 RMSNSLFLVYFGLNHHQDLAHTVCFGRPRELTHEIFNHDGLAEDPSLYLHAPCVTDP 360
301 SMSNSLFLVYFGLNQPHSQLAHHTICFGRPRELTHEIFNHDGLAEDPSLYLHSPCVTDP 360
361 SLAPPCGSGYVYVLAAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVTHRMFT 420
361 SLAPPCASFYVLAAPVPHLGNAPLDWAQEGPKLRDRIFDYLEERYPGLRSQVTHRMFT 420
421 PFDPRDELNAMQGSAPFSEPILOTQSAWFRPNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 ROTSRAHWAILGSLFIBPPSLTQGLF--AANATRH-SNLYLVAAAGTHPGAGIPGVVGLA 477
481 KATAGLMEDL 491
478 ESTASLMIEDL 488

RESULT 7
S-08-096-623A-10
Sequence 10; Application US/08096623A
Patent No. 5684238
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L.
APPLICANT: Brinkhaus, Friedrich L.
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H.
APPLICANT: Yarger, James G.
APPLICANT: Yen, Huel-Chen B.
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1501
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-623A-10

Query Match 70.8%; Score 1845.5; DB 1; Length 489;
Best Local Similarity 72.1%; Pred. No. 1.5e-175;
Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

QY 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDP 60
DB 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQGGFTFDAGPTVITDP 60
QY 61 SAIEELFALAGKQKDYVELLPVTFYELCWESGKVFYNDQALQAEQAOIQFNPRDVAG 120
DB 61 TALEALFTLAGRMEDYVRLLPVTFYELCWESGKTYLDYANDSFELEAQITQFNPRDVAG 120
QY 121 YRAFLDYGRAVFNEGYLKIGTVPFSLFKDMLRAAPQLAKQAWRSVYSKVAGYIEDEHLR 180
DB 121 YRRFLAYSQAVFQEGYLRGLSVFSLFRDMLRAGPQLLKQAWQSVYQSVSRFIEDEHLR 180
QY 181 QAFSFSLLVGNPPTATSSITLIHALRENGVWPPRGCTGALVNGMKLPDGLGGEVVL 240
DB 181 QAFSFSLLVGNPPTATSSITLIHALRENGVWPPRGCTGALVNGMKLPDGLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDDGRRETCVASNADVHTYRDLLSQHPAAKQAKLQSK 300
DB 241 NARVVELVADNRVSVQRLADGRIFEDTAVASNADVHTYKLLGTIPVQKRAARLERK 300
QY 301 RMSNSLFLVYFGLNHHQDLAHTVCFGRPRELTHEIFNHDGLAEDPSLYLHAPCVTDP 360
DB 301 SMSNSLFLVYFGLNQPHSQLAHHTICFGRPRELTHEIFNHDGLAEDPSLYLHSPCVTDP 360
QY 361 SLAPPCGSGYVYVLAAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQVTHRMFT 420
DB 361 SLAPPCASFYVLAAPVPHLGNAPLDWAQEGPKLRDRIFDYLEERYPGLRSQVTHRMFT 420
QY 421 PFDPRDELNAMQGSAPFSEPILOTQSAWFRPNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480


```

db      421  RQTRRHAWIALGSLFTIEPPSLTQGLF--AAATRH-SNLYLVAAGTHPGAGIPGVVGLA 477
ry      481  KATAGLMEIDL 491
       :::|||::|
ib      478  ESTASLMIEDL 488

RESULT 8
S-08-660-645A-5
Sequence 5, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Passamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-660-645A-5

Query Match 61.2%; Score 1595, DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0

ry      1  MKPTVIAGFGGLALAIRLQAAGIPVLLLEQRDKPGRAYVYQEQGTFDAGTPTVDP 60
ib      1  MSSALVIGAGFGGLALAIRLQSLAGATTIVEARDKPGRAYVNDQGHVFDAGTPTVDP 60
ry      61  SAIELPFAIAQKQKDVYELLPTVTPFRLCWESKGVFNVDQAQLEAQIQGNPRDVAG 120
ib      61  DSLRELWALSQQPMERDVTLTPVSPFFYRLTWAGRSFEYVNDDELIRQVASFNPADVQ 120
ry      121 YRAFLDYSRAVNEGYLKLGTVPFLSPKDLRAAPOLAKQAKRSVYVKVAGVIEHLR 180
ib      121 YRFHDYAEVYRGYLLKGTTPFLKQLQMLNAPALRLQATRSVHSWAFIQDPHRL 180
ry      181 QAFSPHSLLVCGNPFATSSIYTLTHALERGVVFPFGGTCALVNGMKLFDQLGGEVYL 240
ib      181 QAFSPHTLLVCGNPFSTSIYALHLERRGGVNFARGGTTNOLVAGKVALFERLGGTLL 240

```

b 1 MSALVIGAGFGLALALRLQAGIATTIVEARDKPGGRAYVNDQGHVFDAGPTVWTD 60
61 SAIELFALAGKQLKDYVELLPVTTFYRLCWSGKVFYNDQALQIAQIQFNPDRVAG 120
61 DLSRELWALSGQPMERDVTLLPVSPFYRLTWADGRSFYVNDDBELIPQVASFNPADV 120
121 YRAFIDYSAVENEGYKLGTVPELSFKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 180
121 YRFDYAEVYREGYKLGTVPELSFKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 180
181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKLFQDLGGEVVL 240
181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGRRETCAVASNADVWHTYRDLLSQHPAAKQAKKLSQ 300
241 NARVTRIDTEGRATGVTLDDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAILNRQ 300
301 RMSNSLFVLYFGLNHHDLAHTVFCGPRYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
301 RMSNSLFVLYFGLNHHDLAHTVFCGPRYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPLGRSOLVTHRMFT 420
361 SLAPGCGSMTHYVLAAPVPHLGRADVDWEAEAPGYAERIPEELERRAIPDLRKHLTVSRIPS 420
421 PFDDELNAMQGSFVSVEPILTQSAWFRPNRDKHIDNLYVAGNTHPGAGIGVIGSA 480
421 PADFSTELSAHSGSAFVSVEPILTQSAWFRPNRDRAPNFYIVAGNTHPGAGIGVIGSA 480
481 KATAGLMLEDL 491
481 KATAQWMLSDL 491

RESULT 10

3-09-546-969-5
Sequence 5, Application US/09546969
Patent No. 6207409
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/546.969
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-546-969-5

Query Match 61.2%; Score 1595; DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

Qy 1 MKPTTVIGAGFGLALALRLQAGIATTIVEARDKPGGRAYVNDQGHVFDAGPTVWTD 60
Db 1 MSALVIGAGFGLALALRLQAGIATTIVEARDKPGGRAYVNDQGHVFDAGPTVWTD 60
Qy 61 SAIELFALAGKQLKDYVELLPVTTFYRLCWSGKVFYNDQALQIAQIQFNPDRVAG 120
Db 61 DLSRELWALSGQPMERDVTLLPVSPFYRLTWADGRSFYVNDDBELIPQVASFNPADV 120
Qy 121 YRAFIDYSAVENEGYKLGTVPELSFKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 180
Db 121 YRFDYAEVYREGYKLGTVPELSFKDMLRAAPOLAKLQAWRSVYSKVAGYIEDEHLR 180
Qy 181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKLFQDLGGEVVL 240
Db 181 QAFPSHLLVGNPFATSIYTLIHALERENGWVPPRGCTGALVNMKLFQDLGGEVVL 240
Qy 241 NARVSHMETVGDKIQAQVLEDGRRETCAVASNADVWHTYRDLLSQHPAAKQAKKLSQ 300
Db 241 NARVTRIDTEGRATGVTLDDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAILNRQ 300
Qy 301 RMSNSLFVLYFGLNHHDLAHTVFCGPRYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
Db 301 RMSNSLFVLYFGLNHHDLAHTVFCGPRYRELTHETFNHDLAEDPSLYLHAPCVTDP 360
Qy 361 SLAPGCGSYVYLAAPVPHLGTANLDWAVEGPRLRDRIPDYLEQHYMPLGRSOLVTHRMFT 420
Db 361 SLAPGCGSMTHYVLAAPVPHLGRADVDWEAEAPGYAERIPEELERRAIPDLRKHLTVSRIPS 420
Qy 421 PFDDELNAMQGSFVSVEPILTQSAWFRPNRDKHIDNLYVAGNTHPGAGIGVIGSA 480
Db 421 PADFSTELSAHSGSAFVSVEPILTQSAWFRPNRDRAPNFYIVAGNTHPGAGIGVIGSA 480
Qy 481 KATAGLMLEDL 491
Db 481 KATAQWMLSDL 491

RESULT 11

US-08-980-832-4
Sequence 4, Application US/08980832B
Patent No. 6291204
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/08/980.832B
CURRENT FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-08-980-832-4

Query Match 61.2%; Score 1595; DB 3; Length 494;
Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVIGAGGGLALAIRLOAGIPVLLLEORDKPGGRAYVVOGQFTFDAGPTVITDP 60
1 MSSAIVIGAGGGLALAIRLOAGIATTIVEARDKPGGRAYVWVNDQGHVDFDAGPTVITDP 60
61 SAIEEPFALAGKOLKDYVELLPVTPYRLCWBSGKVFNTDNOQAQLEAQIQOFPNPDVAG 120
61 DLSRELWALSGQPMERDVTLLPVSPFRLTWADGRSFYVNDDELIRQVASFNPADVDG 120
121 YRAPLDYSRAVFEVGEVILKGTVPFLSFKDMRAAPOLAKLOAWRSVYSKVAGYIEDEHLR 180
121 YRPHDYAEVYREGVILKGTTPFLKQMLNAPALMRLQAYRSVHSMWARFIQDPHLR 180
181 QAFSFSHLLVGGNPFATSSITTLIHALEREGVWVPPRGGTGALVNGMKLFQDLGGEVVL 240
181 QAFSFTLLVGGNPFSSIIYALIHALERGGVWFAKGGTNQLVAGWALFELGGLTLL 240
241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSOHPPAAKQAKLOSK 300
241 NARVTRIDTEGDRATGVTLLDGRQADRVASNGDVWHSYRDLLGHTTRGRRTKAAILNRQ 300
301 RMSNSLFVLYFGLNHHDDLAHETVCGFPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RWSMSLFVLYFGLSKRPENLAHESVIFGPYKGLVNEIFNGPELPDQFSWYLSHSPCVTDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDMAVEGPRDLRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGMSTHYVLAPVPHLGRADVWEAEAPGVAERIFEELEERRAIPDLRKHLLTVSRIFS 420
421 PFDFRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGTGPVIGSA 480
421 PADSTELSAHSGSAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGTGPVIGSA 480
481 KATAGLMLEDL 491
481 KATAQVMSDL 491

RESULT 12

IS-09-547-267-5
Sequence 5, Application US/09547267
Patent No. 6613543
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/560,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-547-267-5
Query Match 61.2%; Score 1595; DB 4; Length 494;
Best Local Similarity 61.9%; Pred. NO. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;
QY 1 MKPTTVIGAGGGLALAIRLOAGIPVLLLEORDKPGGRAYVVOGQFTFDAGPTVITDP 60
DB 1 MSSAIVIGAGGGLALAIRLOAGIATTIVEARDKPGGRAYVWVNDQGHVDFDAGPTVITDP 60
QY 61 SAIEEPFALAGKOLKDYVELLPVTPYRLCWBSGKVFNTDNOQAQLEAQIQOFPNPDVAG 120
DB 61 DLSRELWALSGQPMERDVTLLPVSPFRLTWADGRSFYVNDDELIRQVASFNPADVDG 120
QY 121 YRAPLDYSRAVFEVGEVILKGTVPFLSFKDMRAAPOLAKLOAWRSVYSKVAGYIEDEHLR 180
DB 121 YRPHDYAEVYREGVILKGTTPFLKQMLNAPALMRLQAYRSVHSMWARFIQDPHLR 180
QY 181 QAFSFSHLLVGGNPFATSSITTLIHALEREGVWVPPRGGTGALVNGMKLFQDLGGEVVL 240
DB 181 QAFSFTLLVGGNPFSSIIYALIHALERGGVWFAKGGTNQLVAGWALFELGGLTLL 240
QY 241 NARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSOHPPAAKQAKLOSK 300
DB 241 NARVTRIDTEGDRATGVTLLDGRQADRVASNGDVWHSYRDLLGHTTRGRRTKAAILNRQ 300
QY 301 RMSNSLFVLYFGLNHHDDLAHETVCGFPYRELIIHEIFNHDGLAEDFSLYLHAPCVTDP 360
DB 301 RWSMSLFVLYFGLSKRPENLAHESVIFGPYKGLVNEIFNGPELPDQFSWYLSHSPCVTDP 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLDMAVEGPRDLRIFDYLEQHYMPGLRSQLVTHRMFT 420
DB 361 SLAPEGMSTHYVLAPVPHLGRADVWEAEAPGVAERIFEELEERRAIPDLRKHLLTVSRIFS 420
QY 421 PFDFRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGTGPVIGSA 480
DB 421 PADSTELSAHSGSAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGTGPVIGSA 480
QY 481 KATAGLMLEDL 491
DB 481 KATAQVMSDL 491

RESULT 13

US-09-920-923B-4
Sequence 4, Application US/09920923B
Patent No. 6677134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/125944)
CURRENT APPLICATION NUMBER: US/09/920,923B
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 494
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923B-4

Query Match 61.2%; Score 1595; DB 4; Length 494;

Best Local Similarity 61.9%; Pred. No. 1.6e-150;
Matches 304; Conservative 65; Mismatches 122; Indels 0; Gaps 0;

1 MKPTTVIGAGGGLALAIQLQAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDP 60
1 MSSAIVIGAGGGLALAIQLQAGIATTVIQRKPGGRAYVYQEGFTFDAGPTVITDP 60
61 SAIBELPALAGKQKDYVELLPVTPFFYRLCWESGKVFYNDNDQAQLEAQIQOQFNPRDVAG 120
61 DSLRELWALSQPMERDVTLLPVSPFFYRLTWADGRSEYVNDDELIRQVASFNPADV 120
121 YRAFIDYGRVFNNGYLKGTVPFLSPKMLRAAPQAKLQAMRSVYKVGAYIEDEHLR 180
121 YRRFDYAEVYREGYLKGTTPFLKQLQMLNAPALMRLQAVRSVHSMVAREIQDPLR 180
181 QAFSPHSLVGNPPATSSITLIHALEREGWVFPFGGTGALVNGMKLPQDLGSEVVL 240
181 QAFSPHSLVGNPPATSSITLIHALEREGWVFPFGGTGALVNGMKLPQDLGSEVVL 240
241 NARVSHMETVGEKIQAQVLEDGRFETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300
241 NARVTRIDTEGRATGVTLLDGRQLRADTVASNGDVMSYREDLIGHTRGRKTAAILNRQ 300
301 RMSNSLVLYGLNHHHDLAHHTVCGPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSNSLVLYGLNHHHDLAHHTVCGPRYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
361 SLAPGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDVLEQHYMPGLRSQLVTHRMFT 420
361 SLAPGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDVLEQHYMPGLRSQLVTHRMFT 420
421 PPFDELDNAGWGSAPSVPEILTSQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
421 PPFDELDNAGWGSAPSVPEILTSQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVGSA 480
481 KATAGLMEIDL 491
481 KATAGLMEIDL 491

SULT 14
US-09-934-903-16
Sequence 16, Application US/09934903
Patent No. 6660507
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odom, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. 6660507ton, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: CL1646 US NA
CURRENT APPLICATION NUMBER: US/09/934,903
CURRENT FILING DATE: 2001-08-22
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 16
LENGTH: 511
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF8
US-09-934-903-16

Query Match 27.0%; Score 703.5; DB 4; Length 511;
Best Local Similarity 31.7%; Pred. No. 1.9e-61;
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 2 KPTTVIGAGGGLALAIQLQAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDPS 61
DB 5 KHIIVIGAGGGLALAIQLQAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDPS 64
QY 62 AIELPALAGKQKDYVELLPVTPFFYRLCWESGKVFYNDNDQAQLEAQIQOQFNPRDVAG 121
DB 65 VLDEMFCERSEDLFELPLSPWRLLYDRDIFVY-SDRENRAELQVDFDEGTDG 123
QY 122 RAPLDSYRAVNEGY-----LKLGTVPFLSPKMLRAAPQAKLQAMRSVYKVG 172
DB 124 EFMQERKRRFNALYPCITRYSLSK----SFLSL-DLIKALPWIA---FPKSVFNMLQ 175
QY 173 YIEDPHLQAFSPHSLVGNPPATSSITLIHALEREGWVFPFGGTGALVNGMKLPQ 232
DB 176 YFNOEKMLAFQCKYKLGMSPECPALFTMLPYLEHEGYIYHVGKGLNRITAAQAQVIA 235
QY 233 DLGGVFNARVSEMETVGDKIQAQVLEDGRFETCAVASNADVVHTYRDLLSQHPAAAK 292
DB 236 ENGSIHLSNIESLIIENGAAKGVKLGHAELRGDEVIINADFAHAMTHLVK--PGVLK 293
QY 293 Q-AKKIQSKMSNSLVLYGLNHHHDLAHHTVCGPRYRELHIEIFNHDGLAEDFSL 350
DB 294 KYPENLQKREYSCTFMYLGLDKIYD-LPHHTIVFAKDYTTMIRNIFDNKTLTDFSP 352
QY 351 YLHAPCVTDPSLAPGCGSYVVLAPVPHLGTANLDWAVEGPRLDRIFDVLEQHY-MPGL 409
DB 353 YVQNASASDDSLAPAGKALYLVEMPN-NDSGLDWAQAHQCNVEQVLDLGLARLGLSDI 411
QY 410 RSQVTHRMFTPEPFRDELNAGWGSAPSVPEILTSQSAWFRPHNRDKHIDNLYLVGAGTHP 469
DB 412 RAHICEKIIIPQWTDDEHYKATFSLSHKFSQMLYWRPHNRFEELANCYLVGCGTHP 471
QY 470 GAGIPGVIGSAKATAGLM 487
DB 472 GSGLPVIESARISAKLI 489

RESULT 15
US-09-934-903-18
Sequence 18, Application US/09934903
Patent No. 6660507
GENERAL INFORMATION:
APPLICANT: Koffas, Mattheos
APPLICANT: Odom, J. Martin
APPLICANT: Schenzle, Andreas J.
APPLICANT: No. 6660507ton, Kelley C.
APPLICANT: Tomb, Jean-Francois
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production
FILE REFERENCE: CL1646 US NA
CURRENT APPLICATION NUMBER: US/09/934,903
CURRENT FILING DATE: 2001-08-22
PRIOR FILING DATE: September 1, 2001
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 18
LENGTH: 497
TYPE: PRT
ORGANISM: Methylobionas 16a
FEATURE:
OTHER INFORMATION: Amino acid sequences encoded by ORF9
US-09-934-903-18

Query Match 26.4%; Score 687; DB 4; Length 497;
Best Local Similarity 33.3%; Pred. No. 8e-60;
Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

QY 6 VIGAGGGLALAIQLQAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDPSAIEE 65
DB 11 VIGAGGGLSAISLATAGFSVQLIKRNDKVGKLNITMKDGFDFDLPGLSILTMPIFEA 70

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:26:38 ; Search time 58.1706 Seconds
(without alignments)
2389.754 Million cell updates/sec

File: US-09-941-947A-32

Affect score: 2606

Sequence: 1 MKPTTVIGAGFGGLAIRL.....IPGVIGSAKATAGLMLEDLI 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A Genesep 29Jan04:*

1: Genesep1980s:*

2: Genesep1990s:*

3: Genesep2000s:*

4: Genesep2001s:*

5: Genesep2002s:*

6: Genesep2003as:*

7: Genesep2003bs:*

8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2606	100.0	492	5	AEE22314
2	2606	100.0	492	6	AAO16021
3	2606	100.0	492	6	ABP96688
4	2438	93.6	492	2	AAO7466
5	2438	93.6	492	2	AAW82257
6	2438	93.6	492	2	AAW87889
7	2438	93.6	492	2	AAW99099
8	2438	93.6	492	2	AAW26333
9	2050	78.7	494	6	ABW70123
10	1849.5	71.0	489	2	AAW01122
11	1849.5	71.0	489	2	AAW00173
12	1849.5	71.0	489	2	AAW32472
13	1845.5	70.8	489	2	AAW13985
14	1845.5	70.8	489	2	AAW13984
15	1845.5	70.8	489	2	AAW01123
16	1845.5	70.8	489	2	AAW00174
17	1845.5	70.8	489	2	AAW32475
18	1595	61.2	494	2	AAW00871
19	1592	61.1	494	2	AAW69532
20	1415	54.3	526	2	AAW95897
21	703.5	27.0	511	5	AEE22309
22	703.5	27.0	511	5	ABG61588
23	703.5	27.0	511	5	ABG61588
24	703.5	27.0	511	6	ADA14534
25	687	26.4	497	5	AEE22310

26	687	26.4	497	5	ABG61589	ABG61589	High grow
27	687	26.4	497	5	AAU80333	AAU80333	Methylomo
28	687	26.4	497	5	ADA14538	ADA14538	Methylomo
29	624	23.9	502	6	ABW72613	ABW72613	Staphyloc
30	624	23.9	502	6	ADA14542	ADA14542	Staphyloc
31	598.5	23.0	582	2	ABP97465	ABP97465	Blakeslea
32	582	22.7	499	2	AAE64269	AAE64269	Anabaena
33	582	22.7	499	2	ABW49224	ABW49224	Listeria
34	567	21.8	490	5	ABU32531	ABU32531	Protein e
35	562	21.6	582	2	AAW22499	AAW22499	Phaflia d
36	542	20.8	502	6	ABU43877	ABU43877	Protein e
37	531	20.4	544	5	AAO15518	AAO15518	Agromyces
38	482	18.9	543	5	AAO15521	AAO15521	Micrococc
39	484	18.6	497	6	ADA14544	ADA14544	Staphyloc
40	484	18.6	500	6	ABW72616	ABW72616	Staphyloc
41	482	18.5	497	6	ABU16031	ABU16031	Protein e
42	461.5	17.7	530	6	AAE31690	AAE31690	Rhodococc
43	416.5	16.0	548	4	AAE85728	AAE85728	Enzyme in
44	415	15.9	439	4	AAU34200	AAU34200	Staphyloc
45	410.5	15.8	548	4	AAE76640	AAE76640	Corynebac

ALIGNMENTS

RESULT 1
AEE22314
ID AAE22314 standard; protein; 492 AA.

XX AAE22314;

DT 25-JUL-2002 (first entry)

DE Pantoea stewartii phytoene desaturase (CrtL) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

XX anti-oxidant; steroid; flavour; fragrance; electro-optic application;

XX aquaculture; enzyme; phytoene desaturase; CrtL.

XX Pantoea stewartii.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-0229859P.

XX 01-SEP-2000; 2000US-0229907P.

(DUPO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng O, Dicosimo DJ, Koffas M, Miller ES;

XX Odum JM, Picataggio SK, Rouviere PE,

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35512.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates.

XX Claim 19; Page 141-143; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds. The method comprises a transformed metabolising host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the control of regulatory sequences, and contacting the host cell with carbon substrate to produce a carotenoid compound. The method is useful for producing carotenoid compounds such as antheraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in

C the carotenoid biosynthetic pathway and which metabolise single carbon
C substrates. The carotenoids have potent anti-oxidant properties useful in
C diet, and aquaculture elements. The carotenoids are also useful as
C intermediates in the synthesis of steroids flavours and fragrances and
C compounds for potential electro-optic applications. The present sequence
C is Pantoea stewartii phytoene desaturase (CrtL) enzyme used in the
C invention
X
Q Sequence 492 AA;

Query Match 100.0%; Score 2606; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.4e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDP 60
61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
121 YRAPLDYSRAVFNESGKYLKGTVPFLSKDMLRAAPOLAKLQAWRSVYKVGVIIEDEHLR 180
121 YRAPLDYSRAVFNESGKYLKGTVPFLSKDMLRAAPOLAKLQAWRSVYKVGVIIEDEHLR 180
181 QAFSFHSLLVGGNPFATSSITTLIHALEREWGVWFFPRGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFHSLLVGGNPFATSSITTLIHALEREWGVWFFPRGGTGALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLSSOHPPAAKQKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLSSOHPPAAKQKLOSK 300
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNDGLAEDFSLYLHAPCVITDP 360
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNDGLAEDFSLYLHAPCVITDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
421 PFDFRDELNAWQGSFAFSEVPIITQSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDELNAWQGSFAFSEVPIITQSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 2
AA016021
ID AA016021 standard; protein; 492 AA.
AC AA016021;

20-FEB-2003 (first entry)

Pantoea stewartii phytoene desaturase.

Carotenoid; crt.

Pantoea stewartii.

WO200279395-A2.

10-OCT-2002.

25-JAN-2002; 2002WO-US002124.

26-JAN-2001; 2001US-0264329P.

04-MAY-2001; 2001US-0288984P.

XX (CRGI) CARGILL INC.
XX De Souza ML, Kollmann SR, May CA, Schroeder WA;
XX WP.; 2003-075455/07.
XX N-PSDB; ABT14193.
XX Novel isolated nucleic acid useful e.g. to engineer host cells with the
XX ability to produce particular carotenoids and polypeptides useful in cell
XX -free systems to make particular carotenoids.
XX Claim 26; Page 63-64; 74pp; English.
XX The invention comprises the amino acid and coding sequence of a number of
XX carotenoid (crt)-related proteins. The crt-related DNA and protein
XX sequences of the invention are useful for engineering cells which are
XX able to produce carotenoids. The present amino acid sequence represents a
XX crt-related protein of the invention
XX Sequence 492 AA;

Query Match 100.0%; Score 2606; DB 6; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.4e-254;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDP 60
1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQDKPGGRAYVYQEGFTFDAGPTVITDP 60
61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVAG 120
121 YRAPLDYSRAVFNESGKYLKGTVPFLSKDMLRAAPOLAKLQAWRSVYKVGVIIEDEHLR 180
121 YRAPLDYSRAVFNESGKYLKGTVPFLSKDMLRAAPOLAKLQAWRSVYKVGVIIEDEHLR 180
181 QAFSFHSLLVGGNPFATSSITTLIHALEREWGVWFFPRGGTGALVNGMIKLFQDLGGEVVL 240
181 QAFSFHSLLVGGNPFATSSITTLIHALEREWGVWFFPRGGTGALVNGMIKLFQDLGGEVVL 240
241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLSSOHPPAAKQKLOSK 300
241 NARVSHMETVGDKIQAQVLEDGRRFETCAVASNADVVHTYRDLSSOHPPAAKQKLOSK 300
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNDGLAEDFSLYLHAPCVITDP 360
301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFNDGLAEDFSLYLHAPCVITDP 360
361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPEGCGSYVVLAPVPHLGTANLDNAVEGPRLDRIFDYLEQHYMPGLRSQLVTHRMFT 420
421 PFDFRDELNAWQGSFAFSEVPIITQSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
421 PFDFRDELNAWQGSFAFSEVPIITQSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
481 KATAGLMLEDLI 492
481 KATAGLMLEDLI 492

RESULT 3
ABP96688
ID ABP96688 standard; protein; 492 AA.
XX
AC ABP96688;

03-JUN-2003 (first entry)

Pantoea stewartii phytoene desaturase SEQ ID NO:8.

Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
crtI; crtB; crtZ; phytoene desaturase; enzyme; phytoene; carotenoid.
Pantoea stewartii.

WO2003016503-A2.

27-FEB-2003.

15-AUG-2002; 2002WO-US026647.

15-AUG-2001; 2001US-0312646P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

WPI; 2003-268323/26.

N-PSDB; ACC44762.

Novel nucleic acid molecule isolated from *Pantoea stewartii* encoding a
carotenoid biosynthetic enzyme, useful for regulating carotenoid
biosynthesis in an organism.

Claim 4; Page 63-65; 60pp; English.

The present invention describes *Pantoea stewartii* carotenoid biosynthetic
enzymes (I). More specifically described are the geranylgeranyl
pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
(crtS) and beta-carotene hydroxylase (crtZ) enzymes (I). (See ASP96685 to
ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating
carotenoid biosynthesis in an organism, by over-expressing (I) in an
organism, such that the carotenoid biosynthesis is altered in the
organism. (I) and the genes encoding (I) are useful for converting
phytoene to the carotenoids, for creating recombinant organisms that have
the ability to produce various carotenoid compounds, and also for
enhancing or manipulating carotenoid compounds. (I) can also be used for
producing gene products having enhanced or altered activity

Sequence 492 AA;

Query Match 100.0%; Score 2606; DB 6; Length 492;

Best Local Similarity 100.0%; Pred. No. 4.4e-254;

Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTVITDP 60

1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTVITDP 60

61 SAIEELFALAGKQLKDYVELLPVTPTFYLWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

61 SAIEELFALAGKQLKDYVELLPVTPTFYLWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

121 YRAFLDYSAVNEGVLKLGTVPTFLSFKDMLRAAPQLAKLQAWRSVYSKVAGYIIEDEHLR 180

121 YRAFLDYSAVNEGVLKLGTVPTFLSFKDMLRAAPQLAKLQAWRSVYSKVAGYIIEDEHLR 180

181 QAFSFSLSLVGNPNFATSIYTLIHALLREWGVPFGGTGALVNGMIKLPDLGGVWL 240

181 QAFSFSLSLVGNPNFATSIYTLIHALLREWGVPFGGTGALVNGMIKLPDLGGVWL 240

241 NARVSHMETVGDKIQAQVLEDCRRRPTCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300

241 NARVSHMETVGDKIQAQVLEDCRRRPTCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSNSLFLVYFGLNHHHDLAHTTVCFGRYRELTHIEIFNHDGLAEDFSLYLHAPCVTDP 360

301 RMSNSLFLVYFGLNHHHDLAHTTVCFGRYRELTHIEIFNHDGLAEDFSLYLHAPCVTDP 360

361 SLAPEGCGSYVVLAPVPHLGTANLQNAVEGRRLRDIIDYLEQHYMPGLRSQLYTHWFT 420

361 SLAPEGCGSYVVLAPVPHLGTANLQNAVEGRRLRDIIDYLEQHYMPGLRSQLYTHWFT 420

QY 421 PFDFRDELNAWQGSFSAFSEPIILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
DB 421 PFDFRDELNAWQGSFSAFSEPIILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

QY 481 KATAGIMLEDLI 492

DB 481 KATAGIMLEDLI 492

RESULT 4

AA07466
ID AA07466 standard; protein; 492 AA.

XX
AC AA07466;

XX
24-OCT-2003 (revised)

DT 28-JAN-1991 (first entry)

XX Polypeptide with enzymatic activity for the conversion of phytoene into
DE lycopene.

XX Carotenoid biosynthesis; vitamin A; cancer; food coloring.

XX Pantoea ananatis.

PN BP393690-A.

XX 24-OCT-1990.

PF 20-APR-1990; 90EP-00107493.

XX 21-APR-1989; 89JP-00103078.

PR 05-MAR-1990; 90JP-00053255.

XX (KIRI) KIRIN BEER KK.

XX Misawa N, Kobayashi K, Nakamura K;

XX WPI; 1990-32212/43.

DR N-PSDB; AAQ06296.

XX DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn.

PT of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.

PS Claim 4; Fig 4; 40pp; English.

XX Gene products are useful for the synthesis of carotenoids, useful as food
CC coloring, vitamin A precursor, and possibly in prevention of cancer. See
CC also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 492 AA;

Query Match 93.6%; Score 2438; DB 2; Length 492;

Best Local Similarity 93.5%; Pred. No. 4.3e-237;

Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTVITDP 60

DB 1 MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGTFDAGTVITDP 60

QY 61 SAIEELFALAGKQLKDYVELLPVTPTFYLWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

DB 61 SAIEELFALAGKQLKDYVELLPVTPTFYLWESGKVFYNDQDLQAQLQOQFNPRDVAG 120

QY 121 YRAFLDYSAVNEGVLKLGTVPTFLSFKDMLRAAPQLAKLQAWRSVYSKVAGYIIEDEHLR 180

DB 121 YRAFLDYSAVNEGVLKLGTVPTFLSFKDMLRAAPQLAKLQAWRSVYSKVAGYIIEDEHLR 180

QY 181 QAFSFSLSLVGNPNFATSIYTLIHALLREWGVPFGGTGALVNGMIKLPDLGGVWL 240

DB 181 QAFSFSLSLVGNPNFATSIYTLIHALLREWGVPFGGTGALVNGMIKLPDLGGVWL 240

are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3',-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 492 AA;

Query Match 93.6%; Score 2438; DB 2; Length 492;

Best Local Similarity 93.5%; Pred. No. 4.3e-237;

Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

1 MKPTTVICAGFGGLALAIRLQAAGIPVLLLEQORDKPGRAVYVQEQFTFDAGPTVITDP 60

1 MKPTTVICAGFGGLALAIRLQAAGIPVLLLEQORDKPGRAVYVQEQFTFDAGPTVITDP 60

61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESKVFYNDNDQALQAEQIQQFNPRDVAG 120

61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESKVFYNDNDQALQAEQIQQFNPRDVAG 120

121 YRAFIDYSAVFNQGYLKGTVPFSLPKDMLRAAPQALAKQAWRSVYKVGATYIEDEHLR 180

121 YRAFIDYSAVFNQGYLKGTVPFSLPKDMLRAAPQALAKQAWRSVYKVGATYIEDEHLR 180

181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGSEVL 240

181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGSEVL 240

241 NARVSHMETVGDGKIQAQVLEDDGRFPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300

241 NARVSHMETVGDGKIQAQVLEDDGRFPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLKAPCVTDP 360

301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLKAPCVTDP 360

361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRDLRDRIDFYLEQHYMPGLRSQLVTHRMFT 420

361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRDLRDRIDFYLEQHYMPGLRSQLVTHRMFT 420

421 PFDRELNAMQGSFAFSEVPEPILTSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

421 PFDRELNAMQGSFAFSEVPEPILTSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

481 KATAGLMLEDLI 492

481 KATAGLMLEDLI 492

AAW99099 standard; protein; 492 AA.

AAW99099;

17-OCT-2003 (revised)

14-MAY-1999 (first entry)

Erwinia uredovora crtI protein sequence.

Beta-carotene hydroxylase; crtY; crtB; crtE; crtI; xanthophyll;

metabolite.

Pantoea ananatis.

JPL1046770-A.

23-FEB-1999.

07-AUG-1997; 97JP-00213648.

07-AUG-1997; 97JP-00213648.

(KIRI) KIRIN BREWERY KK.

WPI; 1999-208113/18.

N-PSDB; AAX19119.

Beta-carotin hydroxylase - useful for preparation of xanthophylls and their metabolites.

Disclosure; Page 13-15; 17pp; Japanese.

The present invention describes beta-carotin hydroxylase. Beta-carotene can be used in the preparation of xanthophylls and their metabolites. The present sequence represents an Erwinia uredovora crtI protein sequence from the present invention. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 492 AA;

Query Match 93.6%; Score 2438; DB 2; Length 492;

Best Local Similarity 93.5%; Pred. No. 4.3e-237;

Matches 460; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

1 MKPTTVICAGFGGLALAIRLQAAGIPVLLLEQORDKPGRAVYVQEQFTFDAGPTVITDP 60

1 MKPTTVICAGFGGLALAIRLQAAGIPVLLLEQORDKPGRAVYVQEQFTFDAGPTVITDP 60

61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESKVFYNDNDQALQAEQIQQFNPRDVAG 120

61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESKVFYNDNDQALQAEQIQQFNPRDVAG 120

121 YRAFIDYSAVFNQGYLKGTVPFSLPKDMLRAAPQALAKQAWRSVYKVGATYIEDEHLR 180

121 YRAFIDYSAVFNQGYLKGTVPFSLPKDMLRAAPQALAKQAWRSVYKVGATYIEDEHLR 180

181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGSEVL 240

181 QAFSFSLSLVGNGNPFATSSIIYTLIHALERENGWVFPFGGTGALVNGMKLPQDLGGSEVL 240

241 NARVSHMETVGDGKIQAQVLEDDGRFPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300

241 NARVSHMETVGDGKIQAQVLEDDGRFPETCAVASNADVVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLKAPCVTDP 360

301 RMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHEIFNHDGLAEDFSLYLKAPCVTDP 360

361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRDLRDRIDFYLEQHYMPGLRSQLVTHRMFT 420

361 SLAPEGCGSYVVLAPVPHLGTANLDWAVEGPRDLRDRIDFYLEQHYMPGLRSQLVTHRMFT 420

421 PFDRELNAMQGSFAFSEVPEPILTSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

421 PFDRELNAMQGSFAFSEVPEPILTSAMFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

481 KATAGLMLEDLI 492

481 KATAGLMLEDLI 492

AAW99099 standard; protein; 492 AA.

AAW99099;

17-OCT-2003 (revised)

13-JAN-2000 (first entry)

Erwinia uredovora phytoene desaturase enzyme (PDS).

Transgenic plant; herbicide; resistant; tolerant; phytoene desaturase;

PDS; crtI gene; 4-hydroxy-phenylpyruvate dioxygenase; 4HPPD; banana;

cotton; maize; tomato; vine; weed control; isoxazole herbicide family;

KW

6 VIGAGFGGLAALRQAGIPLVLLLEQDQKPGRAVYVYQGGFTFDAGPTVITDPSAIE 65
6 VIGAGFGGLAALRQAGIPLVLLLEQDQKPGRAVYVYQGGFTFDAGPTVITDPSAIE 65
66 LFAAGKQKDYVELLPVTPFYRLCWESGKVFYNDQACLEAQIOQFNPRDVGAFRL 125
66 LFTQAGKMDYVLLPVHPFYRLCWESGKVFYNDQACLEAQIHTFNPRDVGAFRL 125
126 DYSRAVENGYLKGTVDFPLSPKDMRAAPOLAKQAWRSVYSKVAGYIIDEHLRQAFSP 185
126 DYSRAVENGYLKGTVDFPLSPKDMRAAPOLAKQAWRSVYSKVAGYIIDEHLRQAFSP 185
136 HSLVGGNPPATSSITLTHALEREWGVPFPGGTGALVNGMIKLFQDLGGEVYLNARVS 245
136 HSLVGGNPPATSSITLTHALEREWGVPFPGGTGALVNGMIKLFQDLGGEVYLNARVS 245
246 HMETVGDKIQAQVLEDEGRFTCAVASNADVWYHYRDLSSQHPAAAKQKLSKMSNS 305
246 HMETVGDKIQAQVLEDEGRFTCAVASNADVWYHYRDLSSQHPAAAKQKLSKMSNS 305
306 LFLVYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
306 LFLVYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTDPSLAPE 365
366 GCGSYVVLAPVPHLGTANLMDNVAEGPRLDRIFDYLSQHYMPGLRSQIVTHRFTPPDFR 425
366 GCGSYVVLAPVPHLGTANLMDNVAEGPRLDRIFDYLSQHYMPGLRSQIVTHRFTPPDFR 425
426 DELNAGWSAFSVPILTSQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVISAKATAG 485
426 DELNAGWSAFSVPILTSQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVISAKATAG 485
486 LMELE 491
486 LMELE 491
RESULT 10
AAW01122
AAW01122 standard; protein; 489 AA.
AAW01122;
16-OCT-2003 (revised)
25-MAR-2003 (revised)
11-DEC-1996 (first entry)
Phytoene dehydrogenase-4H.
GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;
pigment; food colourant; chloroplast transit peptide; increase yield;
tobacco ribulose bis-phosphate carboxylase-oxygenase.
Pantoea agglomerans.
US5530188-A.
25-JUN-1996.
21-JUL-1993; 93US-00095726.
02-MAR-1990; 90US-00487613.
18-MAY-1990; 90US-00525551.
03-AUG-1990; 90US-00582674.
28-FEB-1991; 91US-00662921.
30-OCT-1991; 91US-00785566.
(STAD) AMOCO CORP.
Yan HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
Ausich RL;

DR WPI: 1996-308823/31.
DR N-PSDB; AAT40792.
XX Increasing prodn. of total carotenoid(s) in a higher plant - by
PT transforming with vector encoding chloroplast transit peptide operably
PT linked to the Erwinia herbicola lycopene cyclase structural gene.
XX Example 10; Fig 11; 99pp; English.
XX The present sequence is that of the Erwinia herbicola phytoene
CC dehydrogenase-4H which produces lycopene biosynthetically from phytoene
CC through four sequential dehydrogenation reactions. Other enzymes involved
CC in the carotenoid biosynthesis pathway include geranylgeranyl
CC pyrophosphate (AAW01119), phytoene synthase (W01121) and lycopene cyclase
CC (AAW01125). Manipulation of, in part., lycopene cyclase gene
CC (AAT40795), by in frame linkage to the chloroplast transit peptide
CC (AAW01124) of the tobacco ribulose bis-phosphate carboxylase-oxygenase
CC gene can lead to increased production of total carotenoids in the
CC chloroplast of transformed plants as compared to native, non-transformed
CC plants of the same type. Beta-carotene is an effective and apparently
CC harmless food colourant and is also in the pathway for biological
CC synthesis of further C40 carotenoids such as zeaxanthin and zeaxanthin
CC diglucoside. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16
CC -OCT-2003 to standardise OS field)
XX Sequence 489 AA;
SQ Query Match 71.0%; Score 1849.5; DB 2; Length 489;
Best Local Similarity 72.3%; Pred. No. 1.4e-177;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;
QY 1 MKPTTVICAGCGGLAALRQAGIPLVLLLEQDQKPGRAVYVYQGGFTFDAGPTVITD 60
DB 1 MKPTTVICAGCGGLAALRQAGIPLVLLLEQDQKPGRAVYVYQGGFTFDAGPTVITD 60
QY 61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFYNDQACLEAQIOQFNPRDVAG 120
DB 61 TALEALFTLAGRMEDYVLLPVKPFYRLCWESGKTXLDYANDSFELEAQIQFNPRDVAG 120
QY 121 YSAFLDYSAVENGYLKGTVDFPLSPKDMRAAPOLAKQAWRSVYSKVAGYIIDEHLR 180
DB 121 YERFLAYSAQVFEQYRLGVSVPFLSPKDMRAAPOLAKQAWRSVYSKVAGYIIDEHLR 180
QY 181 QAFSFSLLVGGNPPATSSITLTHALEREWGVPFPGGTGALVNGMIKLFQDLGGEVVL 240
DB 181 QAFSFSLLVGGNPPATSSITLTHALEREWGVPFPGGTGALVNGMIKLFQDLGGEVVL 240
QY 241 NARVSHMETVGDKIQAQVLEDEGRFTCAVASNADVWYHYRDLSSQHPAAAKQKLSK 300
DB 241 NARVSHMETVGDKIQAQVLEDEGRFTCAVASNADVWYHYRDLSSQHPAAAKQKLSK 300
QY 301 RMSNSLFLVYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTD 360
DB 301 RMSNSLFLVYFGLNHHDDLAHTVCFGRPYRELIHEIFNHDGLAEDFSLYLHAPCVTD 360
QY 361 SLAPEGCGSYVVLAPVPHLGTANLMDNVAEGPRLDRIFDYLSQHYMPGLRSQIVTHRFT 420
DB 361 SLAPEGCGSYVVLAPVPHLGTANLMDNVAEGPRLDRIFDYLSQHYMPGLRSQIVTHRFT 420
QY 421 PFDPRDELNAGWSAFSVPILTSQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIS 480
DB 421 PFDPRDELNAGWSAFSVPILTSQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIS 480
QY 481 KATAGLMELE 491
DB 481 KATAGLMELE 491
QY 478 ESTASLMIEDL 488
DB 478 ESTASLMIEDL 488
RESULT 11
AAW00173
ID AAW00173 standard; protein; 489 AA.
XX
AC AAW00173;

16-OCT-2003 (revised)	Phytoene dehydrogenase-4H encoded by DNA derived from pARC376.
25-MAR-2003 (revised)	
17-OCT-1996 (first entry)	
E. herbicola, geranylgeranyl pyrophosphate synthase; pARC376;	
GGPP synthase; biosynthesis; carotenoid; lycopene;	
farnesyl pyrophosphate; phytoene; FPP; isopentyl pyrophosphate; IPP;	
tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;	
food colourant; herbicide; norflurazon.	
Pantoea agglomerans.	
US5530189-A.	
25-JUN-1996.	
22-JUL-1993; 93US-00096043.	
02-MAR-1990; 90US-00487613.	
18-MAY-1990; 90US-00525551.	
03-AUG-1990; 90US-00562674.	
28-FEB-1991; 91US-00862921.	
30-OCT-1991; 91US-00785568.	
(STAD) AMOCO CORP.	
Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;	
Ausich RL;	
WPI; 1996-308824/31.	
N-PSDB; AAT37094.	
DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for	
prodn. of lycopene, and to produce transgenic plants resistant to	
norflurazon.	
Example 4; Fig 11; 87pp; English.	
This sequence represents Erwinia herbicola phytoene dehydrogenase4H	
encoded by DNA from the plasmid pARC376. Phytoene dehydrogenase-4H has a	
mol. wt. of approx. 51 kD which corresponds to a minimum size of about	
1400 bp for the structural gene, presuming an average amino acid residue	
of about 109. The native phytoene dehydrogenase-4H gene is located	
between positions 7849 and 6380 of plasmid pARC376. Phytoene	
dehydrogenase-4H is an enzyme which is involved in the biosynthesis of	
carotenoids, esp. lycopene, from the ubiquitous precursor, farnesyl	
pyrophosphate. In E. herbicola, phytoene has been found to be formed	
biosynthetically in a two-step process. The initial step is the	
condensation of farnesyl pyrophosphate (FPP) and isopentyl pyrophosphate	
(IPP) to form GGPP. This reaction is catalysed by GGPP synthase. This	
first step is immediately followed by a tail to tail dimerisation of	
GGPP, catalysed by the enzyme phytoene synthase, to form phytoene.	
Lycopene is produced from phytoene by the catalytic action of phytoene	
dehydrogenase-4H. The genes encoding components of the lycopene	
biosynthesis pathway may be used to transform a host cell for the	
commercial production of lycopene which is used as a food colourant.	
Plants transformed with the phytoene dehydrogenase-4H coding sequence are	
protected from the herbicide norflurazon. (Updated on 25-MAR-2003 to	
correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)	
Sequence 489 AA;	
Query Match 71.0%; Score 1849.5; DB 2; Length 489;	
Best Local Similarity 72.3%; Pred. No. 1.4e-177;	
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2	
QY 1 MKPTTVIGAFGGLALAIQLQAAGFVLLLFQDRKPGRAVYVYQGGFTFDAGFTVITDP 60	
1 MKKTIVIGAFGGLALAIQLQAAGFVLLLFQDRKPGRAVYVYVHDDQGGFTFDAGFTVITDP 60	

Qy	61	SAIEELFALAGKOLADYVVELLPVTFFYELCWESGVFNYNDDQAOLEAQIQCFNPRDVAG	120
Db	61	TALEALFALAGMEDYVRLLPVKFFYELCWESGKTDYANDSFELEAQITQFNPRDVEG	120
Qy	121	YRAFDLYSRVNTNEGTLKLTVPFLSPKDLRAAPOLAKIQAWBSVYKVGACIEDEHLR	180
Db	121	YRFEFLAYSOAVFQEGYLRGSPVFFLSFDMRLRAGPQLLKQAWQSVYOSVSRFIEDEHLR	180
Qy	181	QAFSFSHLLVGGNPFATSIYTLIIHALEREWGVNFPFGGTGALVNGMKILFQDLGGEVVL	240
Db	181	QAFSFSHLLVGGNPPFTTSIIYTLIIHALEREWGVNFPFGGTGALVNGMKILFQDLGGETEL	240
Qy	241	NARVSHMETVGGKIQAVQLEDGRRETCAVASNADVDVHTYRDLISQHPAAAKOAKKLOSK	300
Db	241	NARVEELVVDNRVSOVRLADGRIFDTDAVASNADVDVNTYKLLGTIPVGKRAARLERK	300
Qy	301	RMSNSLFLVLYFGLNHHHDLQAHTTVCFGPVRYRELIERIEFNHDLAEDESLYLHAPCVTDP	360
Db	301	SMNSNSLFLVLYFGLNQHSQLAHTTICFGPVRYRELIDEFTGSALADDESLYLHSPCVTDP	360
Qy	361	SLAPECGGYVLVAPVPHLGTANLQWAVEGPELRDRIEDYLEQHTWPGRLRSOLVTHRMFT	420
Db	361	SLAPPPCASFYVLVAPVPHLGNAPLQWAOEGPKLRDRIEDYLEERTWPGRLRSOLVTHQRTFT	420
Qy	421	PFDFRDELNAWOGSAFSPFPIITQTSANFAPHNRDXHIDNLVLVGAGTHPGAGIPGVIGSA	480
Db	421	RQTSRHAWTALIGSLFIEPSPITQGLF--AAATRH-SNLVLYAAAGTHPGAGIPGVVGLA	477
Qy	481	KATAGLMLEDL	491
Db	478	ESTASLMIEDL	488
RESULT 12			
AAW32472			
ID	AAW32472	standard; protein; 489 AA.	
XX	AAW32472;		
AC	XX		
XX	XX		
DT	17-OCT-2003	{revised}	
DT	25-MAR-2003	{revised}	
DT	15-JAN-1998	(first entry)	
XX	XX		
DE	XX	Erwinia herbicola phytoene dehydrogenase-4H.	
XX	XX		
KW	XX	Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;	
KW	XX	lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;	
KW	XX	yeast; plant; vitamin A; cancer.	
XX	XX		
OS	XX	Partosea agglomerans.	
XX	XX		
FH	XX	Location/Qualifiers	
FT	XX	Misc-difference 2 /note= "Encoded by GAA"	
XX	XX		
XX	XX	US5656472-A.	
XX	XX		
PD	XX	12-AUG-1997.	
XX	XX		
PF	XX	07-JUN-1995; 95US-00473512.	
XX	XX		
PR	XX	02-MAR-1990; 90US-00487613.	
PR	XX	18-MAY-1990; 90US-00525551.	
PR	XX	03-AUG-1990; 90US-00562674.	
PR	XX	28-FEB-1991; 91US-00662921.	
PR	XX	21-JUL-1993; 93US-00095726.	
XX	XX		
PA	XX	{STD } AMOCO CORP.	
XX	XX		
PI	XX	Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;	
PI	XX	Brinkhaus FL;	
XX	XX		
DR	XX	WPI; 1997-414592/38.	

N-PSDB; AAT91544.

DNA encoding Erwinia herbicola lycopene cyclase - for producing recombinant enzyme, and transgenic organisms with increased beta-carotene levels.

Example 10; Fig 11; 102pp; English.

A novel DNA molecule has been isolated which encodes an Erwinia herbicola lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA molecule comprises at least 1125 bp and is present in the plasmids pARC147, pARC1509, pARC1510 and pARC1520. The present sequence represents the amino acid sequence corresponding to the structural gene for phytoene dehydrogenase-4H. The new DNA molecule can be used to produce the recombinant enzyme and transgenic organisms, e.g. yeasts or plants, with increased beta-carotene levels. Beta-carotene is used as a colourant in margarine and butter and as an intermediate for vitamin A, and may prevent cancer. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 17-OCT-2003 to standardise OS field)

Sequence 489 AA;

Query Match 71.0%; Score 1849.5; DB 2; Length 489;
Best Local Similarity 72.3%; Pred. No. 1.4e-177;
Matches 355; Conservative 44; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLALAIQLQAAGIPVLLLEQDKPGRAYVYQGGTFDAGTVIDP 60
1 MKPTTVIGAGFGGLALAIQLQAAGIPVLLLEQDKPGRAYVYQGGTFDAGTVIDP 60

61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFVNDQAOLEAQIOQFNPRDVAG 120
61 TALEALFTLAGRMEDYVRLPVKPYRLCWESGKTLDYANDSFELEAQITQFNPRDVAG 120

121 YRAFLDYSRAVNEGYLKGTPVPLSFKMDLRAAPQLAKQAWRSVYSKVAGYIEDEHLR 180
121 YRFLAYSQAVFQEGYLRGSPVFLSFRDMLRAGPQLLQAWQSVYQSVSRFIEDEHLR 180

181 QAFSPHLLVGNPFTSSITLHAIEREWGVPFGGTGALVNGMKLFQDLGGEVVL 240
181 QAFSPHLLVGNPFTSSITLHAIEREWGVPFGGTGALVNGMKLFQDLGGEVVL 240

241 NAFVSHMETVGDKIQAQVLEDDGRFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300
241 NAFVSHMETVGDKIQAQVLEDDGRFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSGNSLVLYFGLNHHDDQIAHHTVCFGRPYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSGNSLVLYFGLNHHDDQIAHHTVCFGRPYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360

361 SLAPEGCGSVYVLAPEHLGTANLQWAVEGPRLRDRIFDYLEQHYMPGLRSQLVTHRMFT 420
361 SLAPPFCASFYVLAPEHLGNAPLQWAVEGPRLRDRIFDYLEQHYMPGLRSQLVTHRMFT 420

421 PFDRELNAWQGSFVPEIITQSNWFPNHRDKHINLYVGAGTHPGAGIPGVIGSA 480
421 ROTSRAHWAITLGLSFTPEPSPITQGLF--AANATRH-SNLXLYVAAGTHPGAGIPGVIGSA 477

481 KATAGLWLDL 491
478 ESTASLMISDL 488

RESULT 13
ARI3985
D ARI3985 standard; protein; 489 AA.
X C AAR13985;
X C AAR13985;
T 24-OCT-2003 (revised)
T 25-MAR-2003 (revised)
T 26-NOV-1991 (first entry)

Phytoene_dehydrogenase-4H from pARC146D.

GGPP; carotenoid; phytoene; zeaxanthin; lycopene.

Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39168).

WO9113078-A.

05-SEP-1991.

04-MAR-1991; 91WO-US001458.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

(STAD) AMOCO CORP.

Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;
Yen HC;

WPI; 1991-281410/38.

N-PSDB; AAQ13720.

Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA encoding enzymes from Erwinia herbicola.

Disclosure; Fig 15(1-4); 313pp; English.

There are a total of six relevant genes in a 7900 bp region that cause E. coli cells to produce GGPP and the carotenoids phytoene through zeaxanthin diglucoside, which is the final prod. identified in the carotenoid pathway contd. in plasmid pARC376 (contg. a ca. 13 kb chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607 (1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase, phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-carotene hydroxylase, and zeaxanthin glycosylase are represented in CC AAQ13716, AAQ13719, AAQ13722, AAQ13724 and AAQ13726 respectively. Recombinant expression plasmids can be used to produce large amts. of the enzymes and hence large amts. of the carotenoids which they synthesise. Carotenoids are pigments with a variety of applications. CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 24-OCT-2003 to standardise OS field)

Sequence 489 AA;

Query Match 70.8%; Score 1845.5; DB 2; Length 489;
Best Local Similarity 72.1%; Pred. No. 3.5e-177;
Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

1 MKPTTVIGAGFGGLALAIQLQAAGIPVLLLEQDKPGRAYVYQGGTFDAGTVIDP 60
1 MKPTTVIGAGFGGLALAIQLQAAGIPVLLLEQDKPGRAYVYQGGTFDAGTVIDP 60

61 SAIEELFALAGKQKDYVELLPVTPFYRLCWESGKVFVNDQAOLEAQIOQFNPRDVAG 120
61 TALEALFTLAGRMEDYVRLPVKPYRLCWESGKTLDYANDSFELEAQITQFNPRDVAG 120

121 YRAFLDYSRAVNEGYLKGTPVPLSFKMDLRAAPQLAKQAWRSVYSKVAGYIEDEHLR 180
121 YRFLAYSQAVFQEGYLRGSPVFLSFRDMLRAGPQLLQAWQSVYQSVSRFIEDEHLR 180

181 QAFSPHLLVGNPFTSSITLHAIEREWGVPFGGTGALVNGMKLFQDLGGEVVL 240
181 QAFSPHLLVGNPFTSSITLHAIEREWGVPFGGTGALVNGMKLFQDLGGEVVL 240

241 NAFVSHMETVGDKIQAQVLEDDGRFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300
241 NAFVSHMETVGDKIQAQVLEDDGRFETCAVSNADVHTYRDLLSQHPAAAKQAKLQSK 300

301 RMSGNSLVLYFGLNHHDDQIAHHTVCFGRPYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360
301 RMSGNSLVLYFGLNHHDDQIAHHTVCFGRPYRELHIEIFNHDGLAEDFSLYLHAPCVTDP 360

301 SMSNSLVLYFGLNQPSQLAHHTICFGRPRYRELIDEIFTGSALADDFSLYLHSPCVTDP 360
 361 SLAPEGCGSYVILAPVPHLGTANLMDWAVEGPRLRDRIFDYLSCHYMPGLRSOLVTHRMFT 420
 361 SLAPPPCASFVILAPVPHLGNAPLDWAQSGPKLRDRIFDYLEERYMPGLRSOLVTRIFT 420
 421 PFDPRDELNAWQGSAPSVPEILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480
 421 RQTSRHAWIAILGSLFTPEPSLTQGLF--AANATRH-SNLYLVAGTHPGAGIPGVVGLA 477

481 KATAGLMEDL 491

478 ESTASLMIEDL 488

RESULT 14

ARL3984

D AAR13984 standard; protein; 489 AA.

X C AAR13984;

X X 24-OCT-2003 (revised)

X T 25-MAR-2003 (revised)

X T 26-NOV-1991 (first entry)

X X Phytoene dehydrogenase-4H.

X X GGPP; carotenoid; phytoene; zeaxanthin; lycopene.

X X Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39368).

X N WO9113078-A.

X D 05-SEP-1991.

X F 04-MAR-1991; 91WO-US001458.

X R 02-MAR-1990; 90US-00487613.

X R 18-MAY-1990; 90US-00525551.

X R 03-AUG-1990; 90US-00562674.

X R 28-FEB-1991; 91US-00662921.

X X (STAD) AMOCO CORP.

X X Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;

X T Yen HC;

X R WPI; 1991-281410/38.

X R N-PSDB; AAQ13719.

X X Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA

X T encoding enzymes from *Erwinia herbicola*.

X X Disclosure; Fig 11(1-4); 313pp; English.

X X There are a total of six relevant genes in a 7900 bp region that cause E.

X C coli cells to produce GGPP and the carotenoids phytoene through

X C zeaxanthin diglucoside, which is the final prod. identified in the

X C carotenoid pathway contd. in plasmid pARC376 (contg. a ca. 13 kb

X C chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607

X C (1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase,

X C phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-

X C carotene hydroxylase, and zeaxanthin glycosylase are represented in

X C AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726

X C respectively. Recombinant expression plasmids can be used to produce

X C large amts. of the enzymes and hence large amts. of the carotenoids which

X C they synthesize. Carotenoids are pigments with a variety of applications.

X C (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to

X C correct PR field.) (Updated on 24-OCT-2003 to standardise OS field)

X X Sequence 489 AA;

X X Query Match 70.8%; Score 1845.5; DB 2; Length 489;

Best Local Similarity 72.1%; Pred. No.: 3.5e-177;
 Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;

Qy 1 MKETTVIGAGFGGLAIALRLOAAGIPVILLRQDKGCGRAYYQEOGFTDAGPTVITDP 60

Db 1 MEKTVVIGAGFGGLAIALRLOAAGIPVILLRQDKGCGRAYYVHDDQGFDDAGPTVITDP 60

Qy 61 SAIEELFALAGKOLKDYVELLPVTTPYRLCWSSGKVFNVDNDQALQEAQIQOQFNPRDVAG 120

Db 61 TALEAFTLAGRMEDYVRLPVKPFYRLCWSSGKTLDYANDSFELEAQITQFNPRDVAG 120

Qy 121 YRAPLDYSRAVFNQGYLKLGTVPPLSFKDLRAAPOLAKLQAWRSVYSKVAGYIDEHLR 180

Db 121 YRRFLAYSQAVFQGYLRGLGSVPFLSFRDKLRAGPQLKLQAWQSVYQSVRFIDEHLR 180

Qy 181 QAPSFESLLVGGNPPATSSITLILHALERENGWVFRGGTGLVNGMIKLFDLGLGGEVYL 240

Db 181 QAPSFHSLLVGGNPPATSSITLILHALERENGWVFRGGTGLVNGMVKLFTDLGSEIEL 240

Qy 241 NARVSHMETVGDKIQAVQLEDGRPFETCAVNASADVHTYRDLLSQHPAAAKQAKLQSK 300

Db 241 NARVEELVADNEVSQVRLADGRIFDTDAVSNADVNTYKLLGTIPVGQRAARLERK 300

Qy 301 RMSNSLVLYFGLNHHDDLAHETVCGPRYRELTHEIFNHDLAEDESLYLHAPCVTDP 360

Db 301 SMSNSLVLYFGLNQPSQLAHHTICFGRPRYRELIDEIFGSAALDDDFSLYLHSPCVTDP 360

Qy 361 SLAPEGCGSYVILAPVPHLGTANLMDWAVEGPRLRDRIFDYLSCHYMPGLRSOLVTHRMFT 420

Db 361 SLAPPPCASFVILAPVPHLGNAPLDWAQSGPKLRDRIFDYLEERYMPGLRSOLVTRIFT 420

Qy 421 PFDPRDELNAWQGSAPSVPEILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSA 480

Db 421 RQTSRHAWIAILGSLFTPEPSLTQGLF--AANATRH-SNLYLVAGTHPGAGIPGVVGLA 477

Qy 481 KATAGLMEDL 491

Db 478 ESTASLMIEDL 488

RESULT 15

AAW01123

ID AAW01123 standard; protein; 489 AA.

XX AC AAW01123;

XX DT 16-OCT-2003 (revised)

XX DT 25-MAR-2003 (revised)

XX DT 11-DEC-1996 (first entry)

XX DE Phytoene dehydrogenase-4H encoded on pARC146D.

XX KW GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;

XX KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;

XX KW pigment; food colourant; chloroplast transit peptide; increase yield;

XX KW tobacco ribulose bis-phosphate carboxylase-oxygenase.

XX OS Pantoea agglomerans.

XX PW US5530188-A.

XX PD 25-JUN-1996.

XX PF 21-JUL-1993; 93US-00095726.

XX PR 02-MAR-1990; 90US-00487613.

XX PR 18-MAY-1990; 90US-00525551.

XX PR 03-AUG-1990; 90US-00562674.

XX PR 28-FEB-1991; 91US-00662921.

XX PR 30-OCT-1991; 91US-00785566.

XX PA (STAD) AMOCO CORP.

I Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;
I Ausich RL;
C
2 WPI; 1996-308823/31.
3 N-PSDB; AAT40793.

C Increasing prodn. of total carotenoid(s) in a higher plant - by
transforming with vector encoding chloroplast transit peptide operably
linked to the Erwinia herbicola lycopene cyclase structural gene.

Example 12; Col 97-100; 98pp; English.

The present sequence is that of a recombinant Erwinia herbicola phytoene
dehydrogenase-4H (from pARCl46D) which produces lycopene biosynthetically
from phytoene through four sequential dehydrogenation reactions. Other
enzymes involved in the carotenoid biosynthesis pathway include
geranylgeranyl pyrophosphate (AAW01119), phytoene synthase (W01121) and
lycopene cyclase (AAW01125). Manipulation of, in part., lycopene
cyclase gene (AAT40795), by in frame linkage to the chloroplast transit
peptide (AAW01124) of the tobacco ribulose bis-phosphate carboxylase-
oxygenase gene can lead to increased production of total carotenoids in
the chloroplast of transformed plants as compared to native, non-
transformed plants of the same type. Beta-carotene is an effective and
apparently harmless food colourant and is also in the pathway for
biological synthesis of further C40 carotenoids such as zeaxanthin and
zeaxanthin diglucoside. (Updated on 25-MAR-2003 to correct PF field.)
(Updated on 16-OCT-2003 to standardise OS field)

Sequence 489 AA;

Query Match 70.8%; Score 1845.5; DB 2; Length 489;
Best Local Similarity 72.1%; Pred. No. 3.5e-177;
Matches 354; Conservative 45; Mismatches 89; Indels 3; Gaps 2;
1 MKPTVIGAGGGGALAIRLOAAGIPVLLLEQDKPGRAVYVQEGTFDAGPTVITDP 60
1 MEKTVIGAGGGGALAIRLOAAGIPVLLLEQDKPGRAVYVQEGTFDAGPTVITDP 60
61 SAIEELFALAGKQIKDYVELLPVTFYRLCWESGKVFYNDQAEQAQIQFNPDRDVAG 120
61 TALEALFTLAGRMEDVRLLPVKPFYRLCWESGKTLVDYANDSFELEAQITQFNPDRVAG 120
121 YRAFLDYSRVAFNEGILKGTVPFLSKDMRAAPOLAKLOAWRSVYSKVAGYIEDEHLR 180
121 YRRFLAYSQAVFOBGYJELGSPFLSPRDLRAGPOLLKLOAWQSVYQSVRSFIEDEHLR 180
181 QAFSFSHLLVGNPFATSSIVTTLIHALEREWGVFPRGGTGALVNGMKLFDLGGCEVVL 240
181 QAFSFSHLLVGNPFATSSIVTTLIHALEREWGVFPRGGTGALVNGMKLFDLGGCEVVL 240
241 NARVSHMETVGDKIQAQVLEDDGRFPETCAVASNADVVHTYRDLISQHPAAAKQKQSK 300
241 NARVEELVADNRVSVQRLADGRIFDFTDAVASNADVVHTYRDLISQHPAAAKQKQSK 300
301 RMSNSLFVLYFGLNHHDLAHTTVCPPRYELIHEIFNHDDGLAEDFSLYLHAPCVTD 360
301 SMSNSLFVLYFGLNHHDLAHTTVCPPRYELIHEIFNHDDGLAEDFSLYLHAPCVTD 360
361 SLAPEGCGSVYVLAPVPHLGTANLDWAVEGFLRDRIDYLEQHYMPGLRSQVLTTHRMFT 420
361 SLAPPPCASPVYVLAPVPHLGTANLDWAVEGFLRDRIDYLEQHYMPGLRSQVLTTHRMFT 420
421 PFDPRDELNAWQGSASFVEPILTQSAWFRPHNEDKHIDNLYLVGAGTHFGAGIRGVIGSA 480
421 RQTSRHAWIAILGSLFIEPSPSLTQGLF--AANATRH-SNLYLVAGAGTHFGAGIRGVIGSA 477
481 KATAGLMLEDL 491
478 ESTASLMIEDL 488

arch completed: February 29, 2004, 14:44:02
Time : 60.2706 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 31.4106 Seconds
(without alignments)
3837.172 Million cell updates/sec

file: US-09-941-947a-30

effect score: 2021

sequence: 1 MOPHYDLILVAGLANGLIA.....SGRPPVPVFAALQAIMTTR 382

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mbc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	ID	Description
1	2021	100.0	382	2	Q8GCS2
2	1715	84.9	382	2	Q47844
3	1494	73.9	386	2	Q8VJ06
4	899	44.5	385	2	Q93C18
5	804	39.8	394	2	Q9KIX3
6	783.5	38.8	382	2	P94791
7	735.5	36.4	386	2	Q9RLH5
8	637	31.5	434	2	O06756
9	156.5	7.7	498	10	Q9M424
10	152.5	7.5	498	10	Q9FV32
11	150	7.4	524	10	Q9XGX3
12	149	7.4	490	10	Q84VG9
13	149	7.4	504	10	Q9M546
14	148.5	7.3	418	2	P72451
15	146	7.2	504	10	Q8LPP7
16	146	7.2	504	10	Q8GTR2

17	145.5	7.2	498	10	Q9LWA6
18	144.5	7.1	417	16	Q7U994
19	142.5	7.1	502	10	Q9AXL1
20	139.5	6.9	430	10	Q8L8H5
21	139.5	6.9	511	10	Q9FV42
22	138.5	6.9	495	10	Q8S3C3
23	138	6.8	902	2	Q8KIP3
24	137.5	6.8	437	10	Q8S3J4
25	137	6.8	377	2	Q7WT70
26	136.5	6.8	517	10	Q8VWR6
27	136	6.7	540	10	Q8LJ81
28	134.5	6.7	516	10	Q9FV43
29	133	6.6	407	16	Q88C14
30	130.5	6.5	529	10	Q9AXL0
31	130.5	6.5	594	10	Q7XAV8
32	130	6.4	549	16	Q8PCT2
33	129.5	6.4	529	10	Q9AXK9
34	128	6.3	399	16	Q7V6N0
35	127.5	6.3	399	16	O06934
36	127.5	6.3	399	16	Q7TVM3
37	126	6.2	426	16	Q7V508
38	124	6.1	411	2	Q83WZ9
39	123.5	6.1	467	16	Q87ZG1
40	123.5	6.1	498	16	Q8PZ0
41	122.5	6.1	520	16	Q988D6
42	122	6.0	403	16	Q7V123
43	121	6.0	549	16	Q8PGP1
44	120	5.9	525	10	Q8L8H4
45	119.5	5.9	397	16	Q82PA6

ALIGNMENTS

RESULT 1

Q8GCS2 PRELIMINARY; PRT; 382 AA.
 AC Q8GCS2
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Lycopene beta-cyclase.
 GN CRTV.
 OS Pantoea stewartii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=66269;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC B200;
 RA deSouza M.L., Kollmann S.R., Schroeder W.A.;
 RT "Carotenoid Biosynthesis (WO 02/079395 A2).";
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY166713; AAN85598.1; -
 DR GO; GO:0045436; P:lycopene beta cyclase activity; IEA.
 DR GO; GO:0016117; P:carotenoid biosynthesis; IEA.
 DR InterPro: IPR008461; Crtv.
 DR Pfam: PF05897; Crtv; 1
 SQ SEQUENCE 382 AA; 43257 MW; 7E93D4976CSFCCB1 CRC64;

Query Match 100.0%; Score 2021; DB 2; Length 382;

Best Local Similarity 100.0%; Pred. No. 3.8e-160;

Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPHYDLILVAGLANGLIALRLOQHPDMRILLIIRAGPEAGNGHTWSFHEEDLTLMQHR 60

Db 1 MOPHYDLILVAGLANGLIALRLOQHPDMRILLIIRAGPEAGNGHTWSFHEEDLTLMQHR 60

QY 61 WIAPLVHEHNPDYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120

Db 61 WIAPLVHEHNPDYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120

QY 121 SVQLADGRITIHASTVIDGRGYTPDSALRVGFGQAFIGQEWOLSAFHLGLSPFIIMDATVDQ 180

```

121 SVQLADGRIIHAISTVIDGRGYTPDSALRVGFOAFIGQEWQLSAPGLSSPIIMDATVDDQ 180
181 NGYRFVYTLPLSATALLEDTHYIDKXANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
181 NGYRFVYTLPLSATALLEDTHYIDKXANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
241 PITLTGDNROFWOQOQFACSLGRLAGLPHFTTGYSLPLAVALADRLSALDVFSSVHOTI 300
241 PITLTGDNROFWOQOQFACSLGRLAGLPHFTTGYSLPLAVALADRLSALDVFSSVHOTI 300
301 AHFAQQRWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
301 AHFAQQRWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
361 ILGKPPVPVFAALQAIMTTHR 382
361 ILGKPPVPVFAALQAIMTTHR 382

RESULT 2
47844
D Q47844 PRELIMINARY; PRT; 382 AA.
C Q47844;
T 01-NOV-1996 (T-EMBLrel. 01, Created)
T 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Dycopene cyclase.
N CRY.
S Pantoea agglomerans.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=549;
N [1]
P SEQUENCE FROM N.A.
MEDLINE=94236237; PubMed=8180698;
X To K.Y., Lai E.M., Lee L.Y., Lin T.P., Kung C.H., Chen C.L.,
A Chang Y.S., Liu S.T.;
T "Analysis of the gene cluster encoding carotenoid biosynthesis in
T Erwinia herbicola H913."
L Microbiology 140:331-339 (1994).
R EMBL: M90698; AAA21262.1; -.
R PR; S52585; S52585.
R GO: 0045436; P:lycopene beta cyclase activity; IEA.
R GO: 0016117; P:carotenoid biosynthesis; IEA.
R InterPro: IPR008461; CRY.
R Pfam: PF05897; CRY; 1.
Q SEQUENCE 382 AA; 43248 MW; A3A3197C91BB1D64 CRC64;

Query Match 84.9%; Score 1715; DB 2; Length 382;
Best Local Similarity 83.8%; Pred. No. 1.2e-134;
Matches 320; Conservative 31; Mismatches 31; Indels 0; Gaps 0;

Y 1 MOPHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGNHTWGFHEEDLTINQHR 60
b 1 MRPHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGNHTWGFHEEDLTINQHR 60
Y 61 WAPLVVHPDQVRFQRRHNSGYCVTSRHFAGILRQOFGHLLWLTAYSAVHA 120
b 61 WAPLVVHPDQVRFQRRHNSGYCVTSRHFAGILRQOFGHLLWLTAYSAVHA 120
Y 121 SVQLADGRIIHAISTVIDGRGYTPDSALRVGFOAFIGQEWQLSAPGLSSPIIMDATVDDQ 180
b 121 AVRLNNGQVISASAVIDGRGYTPNSLNVGFOAFIGQEWQLSAPGLSSPIIMDATVDDQ 180
Y 181 NGYRFVYTLPLSATALLEDTHYIDKXANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
b 181 NGYRFVYTLPLSATALLEDTHYIDKXANLQAEARONIRDYAARQGWPLQTLREEQGAL 240
Y 241 PITLTGDNROFWOQOQFACSLGRLAGLPHFTTGYSLPLAVALADRLSALDVFSSVHOTI 300
b 241 PITLTGDSATAFWQOQPLACSGRLAGLPHFTTGYSLPLAVALADRLSALDVFSSVHOTI 300

```

```

301 AHFAQQRWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
301 THEFAHERWQOQGFPRMLNRMFLAGPADSRWRVMQRFYGLPEDLIARFYAGKLTVDRLR 360
361 ILGKPPVPVFAALQAIMTTHR 382
361 ILGKPPVPVFAALQAIMTTHR 382

RESULT 3
Q8VUJ6
ID Q8VUJ6 PRELIMINARY; PRT; 386 AA.
AC Q8VUJ6;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DB CRY protein.
GN CRY.
OS Pantoea agglomerans pv. milletiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OK NCBI_TaxID=182454;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamiunten H., Hirata R.;
RT "Isolation and characterization of carotenoid biosynthesis genes from
RT Pantoea agglomerans pv. milletiae Wist 801."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076662; BAB79602.1; -.
DR GO: 0045436; P:lycopene beta cyclase activity; IEA.
DR GO: 0016117; P:carotenoid biosynthesis; IEA.
DR InterPro: IPR008461; CRY.
DR Pfam: PF05897; CRY; 1.
SQ SEQUENCE 386 AA; 43674 MW; 1E7DB42C08447AE7 CRC64;

Query Match 73.9%; Score 1494; DB 2; Length 386;
Best Local Similarity 73.4%; Pred. No. 3.4e-116;
Matches 279; Conservative 36; Mismatches 65; Indels 0; Gaps 0;

Y 3 PHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGNHTWGFHEEDLTINQHRMI 62
b 2 PRYDLILVAGLANGLIALRLQOQPSLAILLIDAEEREPGANHTWGFHEEDLTINQHRMI 61
Y 63 APLVHHWPDQVRFQRRHNSGYCVTSRHFAGILRQOFGHLLWLTAYSAVHA 122
b 62 APLVHHWPDQVRFQRRHNSGYCVTSRHFAGILRQOFGHLLWLTAYSAVHA 121
Y 123 QLADGRIIHAISTVIDGRGYTPDSALRVGFOAFIGQEWQLSAPGLSSPIIMDATVDDQNG 182
b 122 TLDDGRVLESDAVIDGRGYTPDGLRMGFSFVQEWQLSEPHGLTAPIMDATVDDQAG 181
Y 183 YRFVYTLPLSATALLEDTHYIDKXANLQAEARONIRDYAARQGWPLQTLREEQGALPI 242
b 182 YRFVYSLPFSADTLIEDTHYIDNATLEGDRARONIRDYAARQGWPLQTLREEQGALPI 241
Y 243 TLTDGDNROFWOQOQFACSLGRLAGLPHFTTGYSLPLAVALADRLSALDVFSSVHOTIAH 302
b 242 TLTDGVAARFQWRDLPCSGRLAGLPHFTTGYSLPLAVALADRLAQMOTFSETLHATQQ 301
Y 303 FAQRWQOQGFPRMLNRMFLAGPAESRWRVMQRFYGLPEDLIARFYAGKLTVDRLRIL 362
b 302 FASQWQOQGFPRMLNRMFLAGPADQRWQVQRFYGLPEGLIARFYAGKLTVDRLRIL 361
Y 363 SGKPPVPVFAALQAIMTTHR 382
b 362 SGKPPVPVFAALQAIMTPHR 381

RESULT 4
Q93C18
ID Q93C18 PRELIMINARY; PRT; 385 AA.
AC Q93C18;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)

```

```

1 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
2 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
3 Lycopene cyclase.
4 CRTX.
5
6 Xanthobacter sp. (strain Py2).
7 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
8 Hypnobiaceae; Xanthobacter.
9 NCBI_TaxID=78245;
10 [1]_
11 SEQUENCE FROM N.A.
12 STRAIN=Py2;
13 Larsen R.A., Metcalf W.W.;
14 Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
15 EMBL; AF498446; AAL01999.1; -.
16 GO; GO:0045436; F:lycopene beta cyclase activity; IEA.
17 GO; GO:0016117; P:carotenoid biosynthesis; IEA.
18 InterPro; IPR008461; CrtY.
19 Pfam; PF05897; CrtY; 1.
20 SEQUENCE 385 AA; 41442 MW; 8F69C80ECF6509FE CRC64;
21
22 Query Match 44.5%; Score 899; DB 2; Length 385;
23 Best Local Similarity 48.3%; Pred. No. 1.4e-66;
24 Matches 182; Conservative 57; Mismatches 136; Indels 2; Gaps 1;
25
26 6 DLILVAGLANGLIALRLQOQHPDMRILLIAGPEAGNHTWSPFHEEDLTINQHRWIAPL 65
27 2 DIVFVAGLANCLMAARLAQRPGELHMLLEAGSVGNGHNSWCHSDSLTAAQRAFLAPP 61
28
29 66 VVHHPDQVPRRRHNSGYCVTSRHFAGILRQOFGOHLWHTAVSAVHAESVOLA 125
30 62 QSLVAGHGVHPAPSRILKGYATISERRVAEVMNERLCAITNARVAHPDHVLE 121
31
32 126 DGRITIASTVIDRGVTPDSALRVGFQAFICQEWLSAPHGLSSPIIMDATVDOQNGYRF 185
33 122 GGERIDASAVDGRGLASRHLDLGYOTFLGCELEMSRPHGLTRPIIMDARVEQLGYRF 181
34
35 186 VYTLPLSATALLIEDTHYIDKANIQAERARQNIRDYAARQGWPLQTLRLREOQALPITL 245
36 182 VTVEEDDCLLVEDITYADGPDLPFADALRGISAYAAQGWAVDYVRSEDLGLPALG 241
37
38 246 GDNROFWOQOQP--ACSGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHQTIAHF 303
39 242 GDINAFLETSGVAPAGLRAGLFHPTTGYSLPDAMALADSVSALDLSGALSAVRSH 301
40
41 304 AQORWQOQGFRLNMLFLAGPAESRVRWQRYGCLPEDIARFYAGKLTIVDRILS 363
42 302 AAAANNGRFPFLNRLFRADPERRYAILQRFYGLSEDLIARFYADRLTLADKARILS 361
43
44 364 GKPPVPVFAALQAIMT 380
45 362 GRPPVSVFALSLVET 378
46
47 Q9KIX3 PRELIMINARY; PRT; 394 AA.
48
49 01-OCT-2000 (TrEMBLrel. 15, Created)
50 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
51 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
52 Lycopene cyclase.
53 CRTX.
54
55 Bradyrhizobium sp. ORS278.
56 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
57 Bradyrhizobiaceae; Bradyrhizobium.
58 NCBI_TaxID=114615;
59 [1]
60 SEQUENCE FROM N.A.
61 STRAIN=ORS278;
62 MEDLINE=20309720; PubMed=10851005;
63 Hannibal L., Lorcun J., Angles d'Ortoli N., Garcia N.,
64 Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;
65 "Isolation and characterization of the canthaxanthin biosynthesis
66 genes from the photosynthetic bacterium Bradyrhizobium sp. strain
67 ORS278".
68 J. Bacteriol. 182:3850-3853(2000).
69 EMBL; AF218415; AAF78200.1; -.
70 GO; GO:0045436; F:lycopene beta cyclase activity; IEA.
71 GO; GO:0016491; F:oxidoreductase activity; IEA.
72 GO; GO:0016117; P:carotenoid biosynthesis; IEA.
73 GO; GO:0006118; P:electron transport; IEA.
74 InterPro; IPR008461; CrtY.
75 InterPro; IPR000172; GMC_oxred.
76 InterPro; IPR000437; Prok_lipoprot_S.
77 Pfam; PF05897; CrtY; 1.
78 Pfam; PF00732; GMC_oxred N; 1.
79 PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
80 SEQUENCE 394 AA; 43435 MW; 8BDD59304EB194A CRC64;
81
82 Query Match 39.8%; Score 804; DB 2; Length 394;
83 Best Local Similarity 43.9%; Pred. No. 1.2e-58;
84 Matches 168; Conservative 63; Mismatches 138; Indels 14; Gaps 6;
85
86 6 DLILVAGLANGLIALRLQOQHPDMRILLIAGPEAGNHTWSPFHEEDLTINQHRWIAPL 65
87 6 DIVVIGGLAGGLIALRLTDAEPOLRVVIEGSSASIAAGNHTWSPFGTIDSSDQHAWLGR 65
88
89 66 VVHHPDQVPRRRHNSGYCVTSRHFAGILRQOFGOHLWHTAVSAVHAESVOLA 124
90 66 VGRWPGVEVFAEHAIRLSTAYLSMTSTRLEAEVQEPFERILRDATATISAT-ADHYVL 124
91
92 125 ADGRITIASTVIDRGVTPDSALRVGFQAFICQEWLSAPHGLSSPIIMDATVDOQNGYR 184
93 125 EGRTLRAPCVIDARGRPVGLALGFQKFLGVRLAAPHGLDVPVIMDATVAGSDGYR 184
94
95 185 FYTTLPLSATALLIEDTHYIDKANIQAERARQNIRDYAARQGWPLQTLRLREOQALPITL 244
96 185 FYTTLPLDQRLIEDITYSDGGLPEQVLRHQRARVYALAKGQIAETIRAEQGVLPVIL 244
97
98 245 TCDNQRFQWQ--QPOACSLRAGLFHPTTGYSLPLAVALAD---RLSALDVFTSSSVH 297
99 245 AGDPGLSKVSDSPRV--GLAALLVHPTTGYSLPDVAVRADLLTARLAQRGALUSSADR 302
100
101 298 QTIAHFAQORWQOQGFRLNMLFLAGPAESRVRWQRYGCLPEDIARFYAGKLTIVTD 357
102 303 ETIDVGRTIWRGGYVFLNRLFKAAEPSEERILARFYGLDQALIERFYAARIQPD 362
103
104 358 RLRI---LSGKPPVPVFAALQAI 377
105 363 KLRVFMHMLMKPPPIPISSALACL 385
106
107 RESULT 6
108 P94791 PRELIMINARY; PRT; 382 AA.
109
110 AC P94791;
111 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
112 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
113 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
114 DE Lycopene cyclase.
115 GN CRTX.
116 OS Flavobacterium sp. ATCC 21588.
117 OC Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;
118 OC Flavobacteriaceae; Flavobacterium.
119 OX NCBI_TaxID=50286;
120 [1]_
121 RN SEQUENCE FROM N.A.
122 RC STRAIN=R1534;
123 RX MEDLINE=97186694; PubMed=9034310;
124 RA Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,
125 van Loon A.P.;
126 "Isolation and characterization of the carotenoid biosynthesis genes
127 of Flavobacterium sp. strain R1534.";
128 RL Gene 185:35-41(1997).
129 DR EMBL; U62808; AAC44851.1; -.
130 GO; GO:0045436; F:lycopene beta cyclase activity; IEA.

```

R GO: 0016117; P: carotenoid biosynthesis; IEA.
R InterPro: IPR008461; Crty.
R Pfam: PF05897; Crty; 1.
Q SEQUENCE 382 AA; 42369 MW; 6D85452F45D9EF4B CRC64;
Query Match 38.8%; Score 783.5; DB 2; Length 382;
Best Local Similarity 43.7%; Pred. No. 5.7e-57;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
Y 5 YDILVAGLANGLIALRLQQOHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHRWIA- 63
b 3 HDLLIAGAGLALALAVDRRDPARIWMDARSGSPDQHTWSCHDLDLS---PEMLAR 59
Y 64 --PLVVHHPDYQVRPQRSHVNSGYCVTSRHFAGILRQOFGQHLWLTAVSAVHAES 121
b 60 LSEIRGEWTDQEVAFDHSRLTITGYSIEAGAILGLQ---GVDELWNTHTVAFLDTG 116
Y 122 VQIADGRIIHAFTVIDGRGTPDSALRVGFQAFIQEWQLSAPGLSSPIINDATVQDN 181
b 117 ATLTDSRIEACVIDARGAVETPHLTGVGFQKVGVEIETDAPHGVVERPMINDATVPQMD 176
Y 182 GYRFVYTLPLSATALLIEDTHIDKANLQERARQNIQDYAARQGWPLQTLRLREEOGA 241
b 177 GYRFYLLSPFTRILLIEDTRYSDGDLDDGALAQASLDYAARRGWTQOB--MRRERGILP 235
Y 242 ITLTGDNROFWQOQPOAC--SGLRAGLFPHTTGYSLPLAVALADRLSALDVFTSSVHQT 299
b 236 IALAHDAIGFWRDHAGQAVPVGLGAGLFPHTTGYSLPLAVALADRLSALDVFTSSVHQT 294
Y 300 IAHFAQORWQOQGFPMILNRMFLAGPABSRVWQRFYGLPEDIARFYAGKLTVT 359
b 295 VRGWAIDRADRDRLFLNRMFLRGCPPDRYRLIQRFYRLPQPLIERFYAGRLTLADRL 354
Y 360 RLILSGKPPVPVFAALQAI 377
b 355 RIVTGPPPLSQAVRCL 372
RESULT 7
9RLH5 PRELIMINARY; PRT; 386 AA.
C Q9RLH5
T 01-MAY-2000 (T-EMBLrel. 13, Created)
F 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Lycopene beta-cyclase.
N Crty.
S Paracoccus marcusii.
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
C Rhodobacteraceae; Paracoccus.
X NCBI_TaxID=59779;
N SEQUENCE FROM N.A.
P STRAIN=MH1;
A Harker M., Hirschberg J.;
T "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
MH1";
L Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
R EMBL; Y15112; CAB56061.1; --
R GO: 0045436; F: lycopene beta cyclase activity; IEA.
R GO: 0016117; P: carotenoid biosynthesis; IEA.
R InterPro: IPR008461; Crty.
R Pfam: PF05897; Crty; 1.
Q SEQUENCE 386 AA; 42283 MW; 91952F0F33F36800 CRC64;
Query Match 36.4%; Score 735.5; DB 2; Length 386;
Best Local Similarity 41.5%; Pred. No. 5.9e-53;
Matches 158; Conservative 69; Mismatches 139; Indels 15; Gaps 8;
Y 5 YDILVAGLANGLIALRLQQOHPDMRILLIE--AGPEAGNHTWSFHEEDLTINQHRWI 62
b 3 HDVILAGLANGLIALAURAARPDRLVLLDRAAGPSEG--HTWSCHDPLDS---PFWL 57

QY 63 A---PLVVHHPDYQVRPQRSHVNSGYCVTSRHFAGILRQOFGQHLWLTAVSAVHA 119
DB 58 ARKLPRANWPDQEVFPFRRHRLATGYSJGGAALADAVVRSGAETRW--DSDIALIDA 116
QY 120 ESQIADGRIIHAFTVIDGRGTPDSALRVGFQAFIQEWQLSAPGLSSPIINDATVQD 179
DB 117 QGATLSCGTREIAGSVLDGRGPHSRHLTLGPHKFLGVEIETDRPHGVPRPVINDGTVTQ 176
QY 180 QNGYRFVYTLPLSATALLIEDTHIDKANLQERARQNIQDYAARQGWPLQTLRLREEOGA 239
DB 177 RDGYGYIYLLPSPTRILLIEDTRYSDGDLDDALAAASQDVARQGW--TGAEVAREGI 235
QY 240 LPTLTGDNROFWQOQPO--ACSGLRAGLFPHTTGYSLPLAVALADRLSALD--VFTSSSV 296
DB 236 LPTALAHDAAGFWADHAEQFVPVGLRAGFFHVTGYSLPYAAQVADVAGLSPGPTDAL 295
QY 297 HTIAHFAQORWQOQGFPMILNRMFLAGPABSRVWQRFYGLPEDIARFYAGKLTVT 356
DB 296 RGAIRDYAIADRARRDRFLNRMFLRGCPPDRYRLIQRFYRLPQPLIERFYAGRLSVA 355
QY 357 DRLRLILSGKPPVPVFAALQAI 377
DB 356 DQLRIVTGPPPLGTAIRCL 376
RESULT 8
006756 PRELIMINARY; PRT; 434 AA.
ID 006756
AC 006756
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DB Lycopene cyclase.
GN Crty.
OS Erythrobacter longus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Erythrobacter.
OX NCBI_TaxID=1044;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=och 101;
RX MEDLINE=97311406; PubMed=9168123;
RA Matsumura H., Takeyama H., Husakabe E., Burgess J.G., Matsumura T.;
RT "Cloning, sequencing and expression the carotenoid biosynthesis gene,
RT lycopene cyclase and phycoene desaturase, from the aerobic
RT photosynthetic bacterium Erythrobacter longus sp. strain och 101 in
RT Escherichia coli.";
RL Gene 189;169-174 (1997).
RL EMBL; D83513; BAA20275.1; --
DR GO: 0045436; F: lycopene beta cyclase activity; IEA.
DR GO: 0016117; P: carotenoid biosynthesis; IEA.
DR InterPro: IPR008461; Crty.
DR Pfam: PF05897; Crty; 1.
SQ SEQUENCE 434 AA; 48018 MW; 9E440E09BA5AEF47 CRC64;
Query Match 31.5%; Score 637; DB 2; Length 434;
Best Local Similarity 38.9%; Pred. No. 1.1e-44;
Matches 154; Conservative 61; Mismatches 149; Indels 32; Gaps 10;
QY 6 DILVAGLANGLIALRLQQOHPDMRILLIEAGPEAGNHTWSFHEEDLTINQHRWIAPL 65
DB 16 DCAIVGGGLAGGLIALALQARPEFIRVIEAGRTTIGNHRWSWFDSDLSAGRALLADF 75
QY 66 VVHEWF--DYQVRPQRSHVNSGYCVTSRHF--AGILRQOFGQHLWLTAVSAVHAESVQ 123
DB 76 RQTDWEGGVEVPKPYRKLKTAYRSMASDTDFHEGLRALPEGSVILGRKAVGLDARGVD 135
QY 124 LADGRI-----IHASTVIDGRGTPDSALRVGFQAFIQEWQLSAPGLSSPIINDATVD 178
DB 136 LAPSGYQPATRINARSVIDCRSFKPSAHLKGGVFLGRHMLQEPHGVENPVIINDATVD 195
QY 179 Q-----QNG--YRFVYTLPLSATALLIEDTHIDKANLQERARQNIQDYAARQGWPLQTL 232

196 QLAZPHGNGSGSYFVYVPLGSHDVFIEDTYADDPLLDNALSGRIDQVARANGWENGTP 255
 233 LREEQALPITITGDNQWQOQPO---ACSGLEAGLPHPTTGYSLPLAV-----ALADR 284
 256 VHEAGVLPV-LTGGDFSAQDEVRLPGVAIGARGGFTPLTSTYTMCVAVENALAMAEQ 314
 285 LSAIDVFTSSVHTIAHF---AQRWQOQGFPRMLNMLFLAGPABSRWRVMQRFYGLP 341
 315 -----PDLSEQLAAFPDSRRRHSXTGYVLLARFLFPFAAKPRKVKVQRFYGLR 367
 342 SOLIARFYAGLTVTDRLILSGKPPVPFALQAI 377
 368 EGLIERFYAARSNTFDKRVLMGCBPPVAIHSAILAM 403

RESULT 9

Q9M424 PRELIMINARY; PRT; 498 AA.
 DT 01-OCT-2000 (TREMELrel. 15, Created)
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Neoxanthin synthase.
 OS NXS.
 OC Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 CX NCBI_TaxID=41113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20547529; PubMed=11094161;
 RA Al-Babili S., Hugueney P., Schledz M., Walsch R., Frohnmeyer H.,
 Laule O., Beyer P.;
 TI "Identification of a novel gene coding for neoxanthin synthase from
 L Solanum tuberosum";
 FEBS Lett. 485:168-172(2000).
 R EMBL; AJ27136; CAB92977.1; .
 R InterPro; IPR008671; Lycopene_cycl.
 R InterPro; IPR000205; NAD_BS.
 R Pfam; PF05834; Lycopene_Cycl; 1.
 SQ SEQUENCE 498 AA; 56356 MW; 0CAB552D958C6F26 CRC64;

Query Match 7.7%; Score 156.5; DB 10; Length 498;
 Best Local Similarity 22.1%; Pred. No. 0.00016;
 Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;
 Y 5 YDLILVAGLGLIALRLQO---QHPDMRILLIEAGPEA---GGNHTWSFHEEDLTIN-- 57
 D 82 FDVILIGAGPA---GLRLAEHVSKYGIKVCVDPSPLSMFPNNYGVWVDFENGLGDC 137
 Y 58 -QHRWIAPIVWHWPDYQVRF---PQRR-----RHVNSGYCVTSRHFAGILRQOFG 105
 D 138 LDHKW---PMTCVHINDHKTLYGRPYGRVSRKXKLRLINS---CVENR-----VKFYK 186
 Y 106 QHLWHTAVSAVHAE---SVQLADGRIIHAFTVIDGRGYTP-----DSALRWGFOAFTGQ 157
 D 187 AKWV-----KVHEEFESIVCDDGKIRGSLVVDASGFASDFIEYDPRNHGYQIANGV 241
 Y 158 EQWL-SAPHGLSSPIIMD-----ATVQQNGYRFVYVTLPLSATALLIEDTHYID 205
 D 242 LVEVDNHPDLDKWLMDWRDLSHLCNEPFLRVNNAKEFTFLYAMPFDRLVFLBETSLVS 301
 Y 206 KANLOAERARQNIIRYAAQWPLQTLREEQGALPITLTGDNRFQWQOQPOACSGLRAG 265
 D 302 RPLVSYMEVKRWRVARLHGLIKVKSVEIEEKCVPIM---GGPLPRIQNVMAIGG-NSG 357
 Y 266 LFHPTTGYSLPLAVALADRLS 286
 D 358 IVHPSTGYVVAESMALAPVLA 378

RESULT 11

Q9XGX3 PRELIMINARY; PRT; 524 AA.
 DT 01-NOV-1999 (TREMELrel. 12, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Lycopene cyclase.
 OS Citrus paradisi (Grapefruit).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Sapindales; Rutaceae; Citrus.
 CX NCBI_TaxID=37656;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Costa M.C., Moreira C.D., Melton J.R., Otoni W.C., Moore G.A.;
 TI "Developmental expression of carotenoid genes in Citrus.";

RESULT 10
 Q9FV32 PRELIMINARY; PRT; 498 AA.
 AC Q9FV32;
 DT 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Chromoplast-specific lycopene beta-cyclase.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 CX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20461502; PubMed=10995464;
 RA Ronen G., Carmel-Goren L., Zamir D., Hirschberg J.;
 TI "An alternative pathway to beta-carotene formation in plant
 RT chromoplasts discovered by map-based cloning of Beta and old-gold
 RT color mutations in tomato.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:11102-11107(2000).
 DR EMBL; AF254793; AAG21133.1; .
 DR InterPro; IPR008671; Lycopene_cycl.
 DR InterPro; IPR000205; NAD_BS.
 DR Pfam; PF05834; Lycopene_Cycl; 1.
 SQ SEQUENCE 498 AA; 56448 MW; 2C21D3480746D5AA CRC64;
 Query Match 7.5%; Score 152.5; DB 10; Length 498;
 Best Local Similarity 22.1%; Pred. No. 0.00034;
 Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;
 QY 5 YDLILVAGLGLIALRLQO---QHPDMRILLIEAGPEA---GGNHTWSFHEEDLTIN-- 57
 DB 82 FDVILIGAGPA---GLRLAEHVSKYGIKVCVDPSPLSMFPNNYGVWVDFENGLGNC 137
 QY 58 -QHRWIAPIVWHWPDYQVRF---PQRR-----RHVNSGYCVTSRHFAGILRQOFG 105
 DB 138 LDHKW---PMTCVHINDHKTLYGRPYGRVSRKXKLRLINS---CVENR-----VKFYK 186
 QY 106 QHLWHTAVSAVHAE---SVQLADGRIIHAFTVIDGRGYTP-----DSALRWGFOAFTGQ 157
 DB 187 AKWV-----KVHEEFESIVCDDGKIRGSLVVDASGFASDFIEYDPRNHGYQIANGV 241
 QY 158 EQWL-SAPHGLSSPIIMD-----ATVQQNGYRFVYVTLPLSATALLIEDTHYID 205
 DB 242 LVEVDNHPDLDKWLMDWRDLSHLCNEPFLRVNNAKEFTFLYAMPFDRLVFLBETSLVS 301
 QY 206 KANLOAERARQNIIRYAAQWPLQTLREEQGALPITLTGDNRFQWQOQPOACSGLRAG 265
 DB 302 RPLVSYMEVKRWRVARLHGLIKVKSVEIEEKCVPIM---GGPLPRIQNVMAIGG-NSG 357
 QY 266 LFHPTTGYSLPLAVALADRLS 286
 DB 358 IVHPSTGYVVAESMALAPVLA 378

Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DB EMBL: AF152246; AAD38049.2; -
DB InterPro: IPR008671; Lycopene_cycl.
DB InterPro: IPR00205; NAD_BS.
DB Pfam: PF05834; Lycopene_cycl; 1.
SQ SEQUENCE 524 AA; 58553 MW; 51823CF5B39208C CRC64;

Query Match 7.4%; Score 150; DB 10; Length 524;
Best Local Similarity 21.9%; Pred. No. 0.00058;
Matches 88; Conservative 66; Mismatches 162; Indels 86; Gaps 20;

6 DLILVAGLANGLIALRLOOQHDMRILLIEAGPE---ACGNHTW---SFHEEDLTINQH 59
109 DLAVVGGPAG--LAVAGQVSEAGLSVCSIDPSKLIWPNNGYVWVDFEAMDLDCDLD 166
60 RWIAPLVVHEWPDYQVRPQRRRHVNSGYCVTSRHF--AGILRQOQFQHLWLHTA--VSA 116
167 TWSGAVV--HIDD-----NTKDLNRPYGRVNRKLLSKMLQKCIWNGVKTQAVKIV 218
117 VHAESVOLA---DGRTHASTVIDGRG-----YTPDSALRVGFOAFIGQEWQLS 162
219 IHEESKSLIICNDGVTTIQAAVVLDTGFSRCLVQYDKPNPGYQVAYGILAEVEQH--- 274
163 APHGLSSPIIMD-----ATVDOONGY--RFVYTLPLSATALLIEDTHYDKANLQA 211
275 --PFOLDKRVFMDWDSHLNNSQLKANSKIPTFLYAMPSSNRIFLZETSILVARPGVPM 333
212 ERASQNIIDYAAARQGWPLQTLREEQ-----GALPITLTGDNRFQWQOQPOACSGL--R 263
334 KDIOERMVARELKHGLGVKSTEEDEHCVPMGGPLVL-----PQRWVGIGGT 381
264 AGLFHTPTGYSPLAVALADELS-----ALDVTSSSVHQTIAHFAQQRW-----QQQGF 313
382 AGMHPSTGYMVAVTAAPIAVNAIVRSLSORISGRKLSAEVWKDLWPIERRQREF 441

314 FRM-INRMFLAGPAESRWVQRFYGLPDLTIAFYAGKLT 354
442 FCGMDILLKDLPA-----TRFFDAFFDLERHYWGHFLS 477

RESULT 12

ID Q84VG9 PRELIMINARY; PRT; 490 AA.
AC Q84VG9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lycopene beta-cyclase.
EN PSI.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. W22;
RX MEDLINE=22558484; PubMed=12671084;
RA Singh M., Lewis P.E., Hardeman K., Bai L., Rose J.K., Mazourek M.,
RA Chomet P., Brunnell T.P.
RT "Activator Mutagenesis of the Pink scutellum1/viviparous7 Locus of
RT Maize."
RL Plant Cell 15:874-884 (2003).
DR EMBL: AY206862; AA018661.1; -
DR InterPro: IPR00205; NAD_BS.
SQ SEQUENCE 490 AA; 53360 MW; 05456FE5D355D1A4 CRC64;

Query Match 7.4%; Score 149; DB 10; Length 490;
Best Local Similarity 21.6%; Pred. No. 0.00064;
Matches 96; Conservative 69; Mismatches 164; Indels 116; Gaps 25;

6 DLILVAGLANGLIALRLOOQHDMRILLIEAGPEAGNHTWSFHEEDLTINQHMTI--- 62

DB DLAVVGGPAGLAVARQVAE--AGLSVCAIDPSPAV-----VW-----PNNTYGVWVDEF 115
63 APLVVH-----WPDYQVRPQRRRHVNSGYCVTSRHFAGILRQOQFQOHLW----- 109
116 EAMGLSHCLDTVWPSASV-----FIDDGAKSLORPYARVARKLLKSTWDRCVANGV 168
110 -LHTA--VSAVHAESVOLA---DGRTHASTVIDGRGTP-----DSALRVGFOAFIGOE 158
169 VFHQAKVAKAVHYDASSLLICDDGVAVPASVLDATGFSRCLVQYDKPNPGYQVAYGIL 228
159 WQLSA--PHGLSSPIIMD-----ATVDOONGY--RFVYTLPLSATALLIEDTHYDK 206
229 AEVDAHFPDIDKLMFMDWRDHSHPGSEIRERRRRIPTFLYAMPFPTRIFLETSILVAR 288
207 ANLOAESARQNIIDYAAARQGWPLQTLREEQ-----GALPITLTGDNRFQWQOQPOACS 260
289 PGLAMDDIQERMAARLRLHGLIRVRSVEEDERCVPIMGGPLVL-----PQRVV 336
261 GL--RAGLFHTPTGYSPLAVALADELS-----ALDVTSSSVHQTIAHFAQQRW 308
337 GIGTAGKVPSTGYMVAVTAAPIAVNAIVRSLSORISGRKLSAEVWKDLWPIERRQREF 396
309 -----QQQGFPMFLNRMLFLAGPAESRWVQRFYGLPDLTIAFYAGKLT 345
397 PANRRQREFP--FCGMDVLLKLDLGGTRFFDAFFDLERHYWGHFLSRLFLPELLMFL 455
346 ARFYAGKLTVTDRILRLISGKPPVPV 370
456 ALP--GNASNSRLEIMA--KGTVPL 477

RESULT 13
Q9M546
ID Q9M546 PRELIMINARY; PRT; 504 AA.
AC Q9M546;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lycopene beta-cyclase.
OS Citrus sinensis (Sweet orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=2711;
RN [1]
RP SEQUENCE FROM N.A.
RA Xu C.J., Zhang S.L.;
RT "Molecular cloning of lycopene beta-cyclase gene from orange (Citrus
RT sinensis)."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF240787; AAP44700.2; -
DR InterPro: IPR008671; Lycopene_cycl.
DR InterPro: IPR00205; NAD_BS.
DR Pfam: PF05834; Lycopene_cycl; 1.
SQ SEQUENCE 504 AA; 56460 MW; 7C12B2BFB464481C CRC64;

Query Match 7.4%; Score 149; DB 10; Length 504;
Best Local Similarity 21.5%; Pred. No. 0.00067;
Matches 86; Conservative 64; Mismatches 168; Indels 82; Gaps 17;

6 DLILVAGLANGLIALRLOOQHDMRILLIEAGPE---ACGNHTW---SFHEEDLTINQH 60
89 DLAVVGGPAG--LAVAGQVSEAGLSVCSIDPSKLIWPNNGYVWVDFEAMDLDCDLD 146
61 WIAPLVVHEWPDYQVRPQRRRHVNSGY-----CVTSRHFAGILRQOQFQOHLWLHTA 113
147 TWSGAVVHIDDTTKDLDRPYGRVNRKLLSKMLQKCIW-----GVKPHQAKV 195
114 VSAVHAESVOLA---DGRTHASTVIDGRGTP-----DSALRVGFOAFIGQEWQLSA-P 164
196 IKVHIEESKSLIICNDGVTTIQAAVVLDTGFSRCLVQYDKPNPGYQVAYGILAEVEHP 255
165 HGLSSPIIMD-----ATVDOONGY--RFVYTLPLSATALLIEDTHYDKANLQAER 213

```

> 256 FDLQWVNDWRDLSHNNSELKANSKIPTFLYAMPFSSNRIFLESTSLVARPVPMKD 315
> 214 ARQNIIDYAAARQGWPLQTLLEEQ-----GALPITLTGNROFWQOQOACSGL--RAG 265
> 316 IQERVARLKHGKIKVRSIBEDHCVCIPMGGLPVL-----PQVVVGIGTAG 363
> 266 LFHPTGYSPLAVALADRLS-----ALDVTSSSVHQTIAHFAQQRW-----QQQFFR 315
> 364 MVHSTGVTWARTLAAPIVANAIVRSLSRSDRSISGHKLSAEVWKOLWPIERRRQREFFC 423
> 316 M-LNRMLFAGPESRWVQRFYGLPEDIARFYAGKLT 354
> 424 FGMDILLKLDLPA-----TRRFPDAPDLEPRYWHGFLS 457
>
> RESULT 14
> 2451 PRELIMINARY; PRT; 418 AA.
> P72451;
> 01-FEB-1997 (TremBLrel. 02, Created)
> 01-FEB-1997 (TremBLrel. 02, Last sequence update)
> 01-OCT-2003 (TremBLrel. 25, Last annotation update)
> Lycopene cyclase.
> Streptomyces griseus.
> Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
> Streptomycineae; Streptomycetaceae; Streptomyces.
> NCBI_TaxID=1911;
> [1]
> SEQUENCE FROM N.A.
> STRAIN=JA933;
> MEDLINE=97074881; PubMed=8917308;
> Schumann G., Nuerberger H., Sandmann G., Kruegel H.J.;
> "Activation and analysis of silent carotenoid biosynthetic genes from
> Streptomyces griseus.";
> Mol. Gen. Genet. 252:658-666 (1996).
> EMBL; X95596; CA864855.1; -
> GO; GO:0045436; F:lycopene beta cyclase activity; IEA.
> GO; GO:0016117; P:carotenoid biosynthesis; IEA.
> InterPro; IPR008461; Ctry.
> InterPro; IPR000205; NAD_BS.
> Pfam; PF05897; Ctry; 1.
> SEQUENCE 418 AA; 46239 MW; 776988E02FC6FF0 CRC64;
>
> Query Match 7.3%; Score 148.5; DB 2; Length 418;
> Best Local Similarity 23.9%; Pred. No. 0.00057;
> Matches 99; Conservative 59; Mismatches 178; Indels 79; Gaps 24;
>
> 1 MOPHYDLILVAGLANGIALRLQOQHPD--MRILLIEA--GPEAGNHTWSFEEZ--- 52
> 1 MPTDFDVVIVGAGAGAGSLAHLCAPSDVPLSVALVDAPPGLRAPPTWCFTWEPGGP 60
>
> 53 -DLTLNQHRMTAPLVHHPDYQVR-----FPQRRHNVSGYCVTSRHFAGILR 101
>
> 61 YD-----PVLASWPLRVRAADGASTVAQLPRLR-----YKMLRSDAPEALVE 104
>
> 102 QQFCQH---LMLHTAVSNVHES-----VOLADGR--LIHASTVIDGR--CYTDSAL 147
>
> 105 QRGRAPDLCMEATASVRDPSGCGEVLTACGERILVGRVDFSRPRLPPAR 164
>
> 148 RVGFAFGQWQLSAGHGLSP---IIMD-ATVDOQNGYRVFVTTPLSATALLIDTHY 203
>
> 165 TLLQHTFG--WVFRTERPVDGTADLMDFRTQPARGLSFGVFLZDPTHALVEVTEF 222
>
> 204 IDKANLQNERARQNIIDY---AARQGWPLQTLLEEQGALPITLTGNROFWQOQOAC- 259
>
> 223 -SPAPLDTDGVRALRHYTHDVLRLG-PLQ-VTAQEGHVIFMT---DGR-FPHKAGRSVY 275
>
> 260 -SGLRAGUFHTTGYSLPLAVALADRLSALDVTSSSVHQTIAHFAQQRWQQQFFRMLN 318
>
> 276 RIGTAGGATRSTGYTF-AAVQORSRAVA-DQLRSGRPLRVAPYGERA-----RLMD 326
>
> 319 RMLFLA---GPAESRWVQRFYGLPEDIARFYAGKLTVDRLRLILSGKPPVVF 370
```

```

Db 327 AVLLRALSGRVGDGFTHRLFRHIPGERLLSFMGDRSQHEDLLILGRTFWPM 381
>
> RESULT 15
> Q8LPP7 PRELIMINARY; PRT; 504 AA.
> AC Q8LPP7;
> DT 01-OCT-2002 (TremBLrel. 22, Created)
> DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
> DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
> DE Lycopene beta-cyclase.
> OS Citrus sinensis (Sweet orange).
> OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
> OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
> OC eurosids II; Sapindales; Rutaceae; Citrus.
> NCBI_TaxID=2711;
> [1]
> RP SEQUENCE FROM N.A.
> RC STRAIN=cv. Red flesh navel orange;
> RA Xu J., Meng H.J., Deng X.X.;
> RT "Molecular cloning of lycopene beta-cyclase gene from Red flesh navel
> orange by using tail-PCR.";
> RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
> DR EMBL; AY094582; AAM21152.1; -
> DR InterPro; IPR008671; Lycopene_cycl.
> DR InterPro; IPR000205; NAD_BS.
> DR Pfam; PF05834; Lycopene Cycl; 1.
> SQ SEQUENCE 504 AA; 56470 MW; E15495367378B6B4 CRC64;
>
> Query Match 7.2%; Score 146; DB 10; Length 504;
> Best Local Similarity 21.7%; Pred. No. 0.0012;
> Matches 86; Conservative 70; Mismatches 165; Indels 76; Gaps 20;
>
> QY 6 DLILVAGLANGIALRLQOQHPDWRILLIEAGPE---AGGNHTW--SFHEED-LTLNQH 59
> DB 89 DLAVVGGGPAG--LAVAGQVSEAGLSVCSIDPSKLIWPNNYGVWDEFEAMDLDCLOT 146
> QY 60 RWIAPLVHHPDYQVRFPQRRHNVSGYCVTSRHF-AGILRQOFGQHLWLHTA--VSA 116
> DB 147 TWSGAVV--HIDD-----NTKOLDRPYGRVNRKLLKSKMLQKCIITGVKFKHOAKIV 198
> QY 117 VHAESVQLA---DGRITHASTVIDRGVTP-----DSALRVGFAFGQWQLSA-PHGL 167
> DB 199 IHESKSLIICNDGVTIQAAVVLDTGFSRCLVQYDYPNGYQVAYGILAEVEEHFDL 258
> QY 168 SSPIMD-----ATVDOQNGY--RPVYTEPLSATALLIEDTHYIDKANLQNERARQ 216
> DB 259 DKWFMWRDLSHNNSELKANSKIPTFLYAMPFSSNRIFLESTSLVARPVPMKDIOE 318
> QY 217 NRDYAAARQGWPLQTLLEEQ-----GALPITLTGNROFWQOQOACSGL--RAGLFH 268
> DB 319 RWVARLKHGKIKVRSIBEDHCVCIPMGGLPVL-----PQVVVGIGTAGMVH 366
> QY 269 PTTGYSLPLAVALADRLS-----ALDVTSSSVHQTIAHFAQQRW-----QQQFFRM-L 317
> DB 367 PSTGYWARTLAAPIVANAIVRSLSRSDRSISGHKLSAEVWKOLWPIERRRQREFFC 426
> QY 318 NRMFLFAGPESRWVQRFYGLPEDIARFYAGKLT 354
> DB 427 DILLKLDLPA-----TRRFPDAPDLEPRYWHGFLS 457
>
> Search completed: February 29, 2004, 14:50:58
> Job time : 35.4106 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

% protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 6.4382 Seconds
(without alignments)
3089.496 Million cell updates/sec

itle: US-09-941-947A-30

effect score: 2021

equences: 1 MQPHYDLILVAGLANGLIA.....SGKPPVPVFAALQAIMTTHR 382

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	ID	Description
1	1678	83.0	382	1 CRTV_PANAN	P21687 Pantoea ana
2	1148.5	56.8	386	1 CRTV_BRWHE	Q01331 erwina her
3	759.5	37.6	386	1 CRTV_AGRAU	P54974 agrobacteri
4	156	7.7	503	1 CCS_CITSI	Q98640 citrus sine
5	152	7.5	526	1 LCVE_LYCES	P65837 lycopersico
6	148.5	7.3	411	1 LCVB_SYNP7	Q55276 synechococc
7	148.5	7.3	498	1 CCS_CAPAN	Q42435 capsicum an
8	148.5	7.3	501	1 LCVB_ARATH	Q38933 arabidopsis
9	141.5	7.0	503	1 LCVB_NARPS	Q40424 narcissus p
10	140	6.9	500	1 LCVB_TOBAC	Q43578 nicotiana t
11	139	6.9	498	1 LCVB_CAPAN	Q43415 capsicum an
12	133	6.6	500	1 LCVB_LYCES	Q43503 lycopersico
13	129.5	6.4	524	1 LCVB_ARATH	Q38932 arabidopsis
14	113.5	5.6	428	1 FIXC_ECOLI	P31575 escherichia
15	106.5	5.3	410	1 Y801_DIRA	Q91w68 deinococcus
16	101.5	5.0	391	1 UBIF_ECOLI	P75728 escherichia
17	98	4.8	867	1 SPMD_ECOLI	P77468 escherichia
18	97	4.8	610	1 GLMS_THIFE	Q56275 t glucosami
19	95.5	4.7	490	1 MBT3_SCHPO	P78937 schizosacch
20	95.5	4.7	829	1 PQOF_PSEFL	P55174 pseudomonas
21	95	4.7	328	1 TH12_SCHPO	P40998 schizosacch
22	94.5	4.7	1150	1 DP3A_ECOLI	P10443 escherichia
23	94	4.7	456	1 YRDX_RHOSH	Q01856 rhodobacter
24	94	4.7	496	1 MQO_FROHM	Q7V886 prochloroco
25	94	4.7	502	1 MQO_SYNFX	Q7U517 synechococc
26	93.5	4.6	961	1 ATCU_YERPE	Q8ZCA7 yersinia pe
27	93	4.6	580	1 NADB_STRCO	Q9X898 streptomyce
28	92.5	4.6	632	1 ETED_SCHPO	P87111 s probable
29	91	4.5	339	1 LSPI_HUMAN	P33241 homo sapien
30	91	4.5	435	1 FIXC_BRAJA	P10331 bradyrhizob
31	90.5	4.5	1160	1 DP3A_SALTY	P14567 salmonella
32	90	4.5	497	1 MQO_PROMP	Q7VZQ2 prochloroco
33	90	4.5	724	1 GLBI_XANCP	Q8PE48 xanthomonas

ALIGNMENTS

RESULT 1

CRTV_PANAN

ID CRTV_PANAN STANDARD; PRT; 382 AA.

AC P21687;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Lycopen cyclase.

GN CRTV.

OS Pantoea ananas (Erwinia uredovora).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.

OX NCBI_TaxID:553;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=20D3;

RA MEDLINE=91072214; PubMed=2254247;

RA Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,

RA Nakamura K., Harashina K.;

RT "Isolation of the Erwinia uredovora carotenoid biosynthetic pathway

RT by functional analysis of gene products expressed in Escherichia

RT coli.";

RL J. Bacteriol. 172:6704-6712(1990).

CC -I- FUNCTION: Catalyzes the cyclization reaction which converts

CC lycopene to beta-carotene.

CC -I- PATHWAY: Carotenoid biosynthesis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D90087; BAA14126.1; -.

DR PIR; C37802; C37802.

DR InterPro; IPR008461; Crty.

DR Pfam; PF05897; Crty; 1.

KW Carotenoid biosynthesis.

SQ SEQUENCE 382 AA; 43047 MW; 62A94222A9EB6D45 CRC64;

Query Match 83.0%; Score 1678; DB 1; Length 382;

Best Local Similarity 82.2%; Pred. No. 8.7e-132;

Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

Qy 1 MQPHYDLILVAGLANGLIALRLQQCHPDMRILLTEAGPEAGGNHTWGFHEDLTLNQR 60

Db 1 MQPHYDLILVAGLANGLIALRLQQCHPDMRILLTEAGPEAGGNHTWGFHEDLTLNQR 60

Qy 61 WIAPLVHVHPDYQVRFPPRRFVNSGYCVTSRHFAGILRQOFGOHLWLHTAVSAVHAE 120

Db 61 WIAPLVHVHPDYQVRFPPRRFVNSGYCVTSRHFAGILRQOFGOHLWLHTAVSAVHAE 120

Qy 121 SVGLADGRILHASTWIDRGYTPDSALRVGFQAFICQEWQLSAPHCLSPIIMDAIVDQO 180

Db 121 SVGLADGRILHASTWIDRGYTPDSALRVGFQAFICQEWQLSAPHCLSPIIMDAIVDQO 180

b 121 SVPLKQGVIGARVIGRGAANASLSVGFQAFIQEWRLSHPHGLSSPIIMDATVDQ 180
 y 181 NGRVFYVPLSNTALLIEDTHYIDKANLQAEARONIRDYAARGWPLQTLRREOGAL 240
 b 181 NGRVFYVPLSNTALLIEDTHYIDKANLQAEARONIRDYAARGWPLQTLRREOGAL 240
 y 241 PITLGTGNQFQWQOQPCAGSLRAGLFPHFTTGYSLPLAVALADRLSALDVFTSSVHOTI 300
 b 241 PITLGTGNQFQWQOQPCAGSLRAGLFPHFTTGYSLPLAVALADRLSALDVFTSSVHOTI 300
 y 301 AHAQORWQOQGFRLMRLMFLAGPAESRWVQSFYGLPEDIARFYAGKLTVDRLRILSS 360
 b 301 THEARWQOQGFRLMRLMFLAGPAESRWVQSFYGLPEDIARFYAGKLTVDRLRILSS 360
 y 361 ILSGKPPVPVFAALQAIMTTHR 382
 b 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 2
 RTY ERWHE STANDARD; PRT; 386 AA.
 C Q01331;
 T 01-APR-1993 (Rel. 25, Created)
 T 01-APR-1993 (Rel. 25, Last sequence update)
 T 10-OCT-2003 (Rel. 42, Last annotation update)
 E Lycopene cyclase.
 N Crty.
 S Erwinia herbicola.
 C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 C Enterobacteriaceae; Pantoea.
 X NCBI_TaxID=549;
 N [1]
 P SEQUENCE FROM N.A.
 C STRAIN=ERO10;
 X MEDLINE=931138098; PubMed=8422926;
 A Hundle B.S., O'Brien D.A., Beyer P., Kleinig H., Hearst J.E.;
 T "In vitro expression and activity of lycopene cyclase and
 T beta-carotene hydroxylase from Erwinia herbicola.";
 L FEBS Lett. 315:329-334(1993).
 C -!- FUNCTION: Catalyzes the cyclization reaction which converts
 C lycopene to beta-carotene.
 C -!- PATHWAY: Carotenoid biosynthesis.
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C
 C EMBL; M87280; AAA64980.1; -;
 R PIR; S52981; S52981.
 R InterPro; IPR008461; Crty.
 R Pfam; PF05897; Crty; 1.
 R Carotenoid biosynthesis.
 K Carotenoid biosynthesis.
 Q SEQUENCE 386 AA; 43341 MW; 54A40563BFCFA980 CRC64;

Query Match 56.8%; Score 1148.5; DB 1; Length 386;
 Best Local Similarity 58.3%; Pred. No. 7.4e-88;
 Matches 216; Conservative 53; Mismatches 102; Indels 1; Gaps 1;
 y 6 DLILVGLAGLIALRLQOQHPDMRILLIEAGPEAGNHTWSFHEEDLTLMQHRWIAPL 65
 b 3 DLILVGLAGLIALRLQOQHPDMRILLIEAGPEAGNHTWSFHEEDLTLMQHRWIAPL 62
 y 66 VVHEWPDYQVRFPORRHVNSGYCYTSRHFAGILRQCFQGHLLHTAVSAVHA 125
 b 63 VAHAWPGYQVFDLRRRLARGYVTSRHFAGILRQCFQGHLLHTAVSAVHA 122
 y 126 DGRIRHASTVIDGRGYTPDSALRVGFQAFIQEWRLSHPHGLSSPIIMDATVDQ 185

Db 123 NGEALLAGAVIDRGVTVASSAMQTYQLFLGQQWRLLTQPHGLTVPILMDATVAQOQGYRF 182
 Qy 186 VYTLPLSATALLIEDTHYIDKANLQAEARONIRDYAARGWPLQTLRREOGALPITLT 245
 Db 183 VYTLPLSATALLIEDTHYIDKANLQAEARONIRDYAARGWPLQTLRREOGALPITLT 242
 Qy 246 GDNRFQWQOQPCAGSLRAGLFPHFTTGYSLPLAVALADRLSALDVFTSSVHOTI 304
 Db 243 GDIQALWADAPGVPRSGMRAGLFPHFTTGYSLPLAVALADRLSALDVFTSSVHOTI 302
 Qy 305 QQWQOQGFRLMRLMFLAGPAESRWVQSFYGLPEDIARFYAGKLTVDRLRILSS 364
 Db 303 ERWRWQOQGFRLMRLMFLAGPAESRWVQSFYGLPEDIARFYAGKLTVDRLRILSS 362
 Qy 365 KPPVPVFAALQAIM 378
 Db 363 KPPVPVFAALQAIM 376

RESULT 3
 ID CRTY AGRU STANDARD; PRT; 386 AA.
 AC P54974;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lycopene cyclase.
 GN Crty.
 OS Agrobacterium aurantiacum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OX NCBI_TaxID=44155;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96062243; PubMed=7592436;
 RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiura S., Saito T.,
 RA Ohtani T., Miki W.;
 RT "Structure and functional analysis of a marine bacterial carotenoid
 RT biosynthesis gene cluster and astaxanthin biosynthetic pathway
 RT proposed at the gene level.";
 RL J. Bacteriol. 177:6575-6584(1995).
 C -!- FUNCTION: Catalyzes the cyclization reaction which converts
 C lycopene to beta-carotene.
 C -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
 C biosynthetic pathway.
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C
 C EMBL; D58420; BAA09593.1; -;
 DR InterPro; IPR008461; Crty.
 DR Pfam; PF05897; Crty; 1.
 KW Carotenoid biosynthesis.
 SQ SEQUENCE 386 AA; 42202 MW; B14C41B688AF78AC CRC64;

Query Match 37.6%; Score 759.5; DB 1; Length 386;
 Best Local Similarity 43.0%; Pred. No. 1.4e-55;
 Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;
 Qy 5 YDILVGLAGLIALRLQOQHPDMRILLIEAGPEAGNHTWSFHEEDLTLMQHRWI 62
 Db 3 HDVLVGLAGLIALRLQOQHPDMRILLIEAGPEAGNHTWSFHEEDLTLMQHRWI 57
 Qy 63 A---PLVHEWPDYQVRFPORRHVNSGYCYTSRHFAGILRQCFQGHLLHTAVSAVHA 119
 Db 58 ARUKLFRANWPDQVRFPORRHVNSGYCYTSRHFAGILRQCFQGHLLHTAVSAVHA 116
 Qy 120 ESQVLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWRLSHPHGLSSPIIMDATVDQ 179

```

b 117 QGATLSCGTEIAGAVLDGEGAPSRHLTVGQKFGVGTETDRDAGVPRFMKMDATVQ 176
y 180 QNGYRVYTLPLSATALLIEDTHYIDKAMQAEARQNRDVAARQGMPLQTLLEBOQA 239
b 177 QDGYRFIYLLPFSPTRIILEDTRYSGDGLDDALAAASHDYARQQGM-TGAERVREGI 235
y 240 LPTLTGDNRRQFQQQP--QACSGLAGLPHPTTGYSLPLAVALADRLSALO-VTSSSV 296
b 236 LPTALAHDAAGFWADHAGGVFVGLRAGFHPTVGTSLFYAAQVADVAGLSGPGTDL 295
y 297 HOTIAHFAQORWQQOQFFRMLNRMFLAGPAESRWVQFYLCPEDLIARFYAGKLTVT 356
b 296 RGAIRDYADRDRDRFLNRMFLRGCAPDRRYTLQRFYRMPHGLIERFYAGRLSVA 355
y 357 DRLGILSGKPPVFAALQAI 377
b 356 DQURIVTGKPPPLGTAIRCL 376

RESULT 4
CS CITSI
D_CCS CITSI STANDARD; PRT; 503 AA.
C Q9S3A0;
I 16-OCT-2001 (Rel. 40, Created)
I 16-OCT-2001 (Rel. 40, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
E Capsanthin/capsorubin synthase, chloroplast precursor.
N CCS.
S Citrus sinensis (Sweet orange).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
C eurosids II; Sapindales; Rutaceae; Citrus.
X NCBI_TaxID=2711;
N [1]
P SEQUENCE FROM N.A.
T Xu C.J., Chen D.M., Zhang S.L.;
T "Molecular cloning of capsanthin/capsorubin synthase gene from orange
T (Citrus sinensis).";
L Submitted (JUL-1999) to the EMBL/GenBank/DBSJ databases.
C -/- FUNCTION: CATALYZES THE CONVERSION OF THE UBIQUITOUS 5,6-
C EPOXYCAPSOTENIDS, ANTHEPAXANTHIN AND VIOLAXANTHIN, INTO CAPSANTHIN
C AND CAPSORUBIN, RESPECTIVELY.
C -/- PATHWAY: Carotenoid biosynthesis.
C -/- SUBCELLULAR LOCATION: Chloroplast; chromoplast (By similarity).
C -/- SIMILARITY: Belongs to the lycopene cyclase family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C
C EMBL; AF169241; AAF16189.1;
C InterPro: IPR008671; Lycopene cycl.
C Pfam; PF05834; Lycopene_cycl_1.
C Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
C TRANSIT ? CHLOROPLAST (POTENTIAL).
C CHAIN ? 503 CAPSANTHIN/CAPSORUBIN SYNTHASE.
C NP BIND 88 117 NAD (POTENTIAL).
C SEQUENCE 503 AA; 56663 MW; 2C957F0AA91075E6 CRC64;

Query Match 7.7%; Score 156; DB 1; Length 503;
Best Local Similarity 21.2%; Pred. No. 2.2e-05;
Matches 94; Conservative 69; Mismatches 166; Indels 114; Gaps 23;

y 5 YDLILVAGLGLALRLQQOHPDWRLILLIAGPEA---GGNHTWSFHEEDLTINQHRW 61
> 86 YDVIITGCGAGLRLAEQVSSRH-SVKVCVDVPSLSTWPNNGYGVWDEFEDIGL----- 139

```

```

Qy 62 IAPLVVHWP-----DYQVRFPPRRHVNVCYCVTSRHF--AGILRQFQOHLKLT 112
Db 140 -VDCDKTWPMTCVFINDHKTKYDRP-----YGRVSRNLIKTKLENCVNSGVFKHK 191
Qy 113 AV--SAVHAB--SVOLADGRIIHAIVTIDGRGYTP-----DSALRVGFQAFITGOEWQL- 161
Db 192 AKVMVNHQDFESSIVCDGNEIKASLIVDASGFASFVEYDKPRNHGQIAHGILAEVE 251
Qy 162 SAPHGLSPPIIMDATVQQNGYR-----FVYTLPLSATALLIEDTHYIDKANL 209
Db 252 SHFPFDLDKMWLDWR--DSHLGNEFYLRASNLKLPFLYAMPFDSNLVLEETLSVSRPVL 310
Qy 210 QASRARONRDVAARQGMPLQTLLEBEQ-----GALPITLTGDNRRQFQQOQQA--CSG 261
Db 311 SYKEVKSRAARLRHNGIRVKRVIEDEKCLIPWGGPLVI-----PQSWAIG 358
Qy 262 LRAGLFHPTGY-----SLPLAVMALDR-----LSALDVFTSSSVHQTIAHFAQORWQQOQF 313
Db 359 GTSGLIHPATGYMVAFTMALAPALADAIAECLGSTRMIRGRLHQKV-----W--NGL 409
Qy 314 F--RMLNRMFLAGPAESRWV---MORFYGLPEDLIARFVAGKLTVDRLRLISG-- 364
Db 410 WPIDRRCNREFYSFG-METILLKLDLXGTRTFDAPFDLNPYYWHGFLSSRLSLAELAGLS 468
Qy 365 -----KPPVPV 370
Db 469 LSLFGHASNSSRLDIIVTKCPVPL 491

RESULT 5
LCYE LYCES STANDARD; PRT; 526 AA.
AC O65837;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lycopene epsilon cyclase, chloroplast precursor (EC 1.14.-.-).
GN CRTL-E-1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. VF36; TISSUE=Leaf;
RA Ronen G., Cohen M., Zamir D., Hirschberg J.;
RT "Regulation of expression of the gene for lycopene epsilon cyclase
RT during fruit ripening of tomato."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBSJ databases.
CC -/- FUNCTION: CATALYZES THE SINGLE CYCLIZATION REACTION WHICH CONVERTS
CC LYCOPENE TO DELTA-CAROTENE AND NEUROSPORENE TO ALPHA-ZEAXANTHIN.
CC -/- PATHWAY: Carotenoid biosynthesis.
CC -/- SUBCELLULAR LOCATION: Chloroplast.
CC -/- SIMILARITY: Belongs to the lycopene cyclase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y14387; CAA74745.1;
CC PIR; T07082; T07082.
CC InterPro: IPR001327; FAD pyr_redox.
CC InterPro: IPR008671; Lycopene cycl.
CC InterPro: IPR001100; Pyr_redox.
CC Pfam; PF05834; Lycopene_cycl_1.
CC PRINTS; PR00368; FADPNE.
CC PRINTS; PR00411; FNDADTASEI.
KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;

```



```
RL Nature 408:820-822(2000).
CC -!- FUNCTION: Catalyzes the double cyclization reaction which converts
CC lycopene to beta-carotene and neurosporene to beta-zeacarotene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the lycopene cyclase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U50739; AAB53337.1; -
DR EMBL: L40176; AAP81880.1; -
DR EMBL: AF117256; AAF82388.1; -
DR EMBL: AAC09400; AAF02819.1; -
DR InterPro: IPR008671; Lycopene_cycl.
DR InterPro: IPR00205; NAD_BS.
DR Pfam: PF05834; Lycopene Cycl; 1.
KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 80 CHLOROPLAST (POTENTIAL).
FT CHAIN 81 501 LYCOPENE BETA CYCLASE.
FT NP_BIND 85 113 NAD (POTENTIAL).
FT CONFLICT 31 31 H -> P (IN REF. 2).
FT CONFLICT 243 243 V -> I (IN REF. 2).
SQ SEQUENCE 501 AA; 56176 MW; C3014578D0BDC4E2 CRC64;
Query Match 7.3%; Score 148.5; DB 1; Length 501;
Best Local Similarity 21.0%; Pred. No. 9.3e-05;
Matches 92; Conservative 66; Mismatches 175; Indels 105; Gaps 20;
QY 6 DLILVAGLGLIALRLQQQHPDRILLIEAGPE--AGGNFW--SFHEED-LTLNQH 59
DB 84 DLIAVGGGPG--LAVAQVSEAGLSVCSIDPSKLIWNNYGVWVDFEAMDLDCLD 141
QY 60 RWIAPLVHHPDYQVRFPPRRRHVNSGYCVTSRHFAGILRQOF-----GQHL 108
DB 142 TWSGAVV-----YVDEGKDLSPYGRVNRKQKSKMLQKCIITNGVKF 185
QY 109 WLHTAVSAVHAB--SVQLADGRIIHAFTVIDRGYTP-----DSALRVGFOAFIGEMQ 160
DB 186 HOSKTVNVHEANSTVCSGVKIQASVLDATGFSRCLVQYDKPNPGYQVAYGIVAE 245
QY 161 LSA-PHGLSSPIIMDATVDQNGY-----RFVYTLPLSATALLIEDTHYIDKAN 208
DB 246 VDGHPFDVDMVPMWDRDKHLDSYELKERNSKIPTFLYAMPFSSNRIFLEETSLVAPG 305
QY 209 LQAEARQNIRDYAARQGWPLQTLRREQ-----GALPITLTGDNQFPWQQOQACSL 262
DB 306 LEMEDIQERMAARLKHGLNVRKIEEDERCVPIMGGLPVL-----PQVVGI 353
QY 263 --RAGLFPTTCYSPLAVADRLS-----ALDVFSSSVH--QTIAHFAQQRW----- 308
DB 354 GCTAGVWHPSTGYVARTLAAPIVANAIVYLGSPSSNLGRQDLSAEVNRDLWPIERR 413
QY 309 QOQGFRLNRMFLAGPAESRWVQRPYGLPED-----LIARFYAGKLTV----- 355
DB 414 ROREFF-CFGMDILKLDLDRFRFFDAFFLQPHYWEGLSSRLFLPELLVFLGLSPSH 472
QY 356 ---TDLRLILSKPPVPV 370
DB 473 ASNTSRLLEMT-KGVFL 489.
RESULT 9
LCYB NARPS
ID LCYB NARPS STANDARD; PRT; 503 AA.
AC Q40424;
DT 16-OCT-2001 (Rel. 40, Created)
```

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lycopene beta cyclase, chloroplast precursor [EC 1.14.-.-].
GN LCYI OR LYC.
OS Narcissus pseudonarcissus (Daffodil).
OC Sukiyota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
OC Narcissus.
OX NCBI_TaxID=39639;
RN (1) TaxID=39639;
RN SEQUENCE FROM N.A.
RC TISSUE:Paracorrilla;
RA Al-Babili S., Hobeika E., Beyer P.;
RA "A cDNA encoding lycopene cyclase from Narcissus pseudonarcissus L.";
RL (in) Plant Gene Register PGR96-107.
RN (2)
RN SUBCELLULAR LOCATION.
RX MEDLINE=97433278; PubMed=9288918;
RX Bonk M., Hoffmann B., von Lintig J., Schledz M., Al-Babili S.,
RA Hobeika E., Kleinig H., Beyer P.;
RA "Chloroplast import of four carotenoid biosynthetic enzymes in vitro
RT reveals differential fates prior to membrane binding and oligomeric
RT assembly.";
RL Eur. J. Biochem. 247:942-950(1997).
CC -!- FUNCTION: Catalyzes the double cyclization reaction which converts
CC lycopene to beta-carotene and neurosporene to beta-zeacarotene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast. Exists as an
CC inactive soluble form and an active membrane-bound form
CC (Probable).
CC -!- SIMILARITY: Belongs to the lycopene cyclase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X98796; CA467331.1; -
DR InterPro: IPR008671; Lycopene_cycl.
DR Pfam: PF05834; Lycopene_cycl; 1.
KW Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
KW Transit peptide; Membrane.
FT TRANSIT 1 85 CHLOROPLAST (POTENTIAL).
FT CHAIN 86 503 LYCOPENE BETA CYCLASE.
FT NP_BIND 90 117 NAD (POTENTIAL).
FT SEQUENCE 503 AA; 56910 MW; 3FDIE355EF184D99 CRC64;
Query Match 7.0%; Score 141.5; DB 1; Length 503;
Best Local Similarity 20.4%; Pred. No. 0.00036;
Matches 85; Conservative 66; Mismatches 150; Indels 115; Gaps 19;
QY 6 DLILVAG-----LANGLIARLQQQHPDRILLIEAGPEAGNHTWSFHEEDLTL 56
DB 89 DLIAVGGGPGARSCSTSGGLSVVSD---PNPKLI---WPNYGVWVDFEOMDL 139
QY 57 ---NQHRTAPLVHHPDYQVRFPPRRRHVNSGYCVTSRHFAGILRQOF----- 104
DB 140 LCLDLATWSGAV-----YVDRSTKMLSRPYARVNRKMLKSKWKKCV 183
QY 105 ---GQHLWLHTAVSAVHAB--SVQLADGRIIHAFTVIDRGYTP-----DSALRVGFOAF 154
DB 184 SNGVRFHQATVYKAMHHEEKSYLICSDGVTIDARVLDATGFSRCLVQYDKPNPGYQVA 243
QY 155 IGOEWQLSA-PHGLSSPIIMD-----ATVQONGY--RFVYTLPLSATALLIEDTH 202
DB 244 YGILAEVHEHPDVDMVPMWDRDKHLNKGKAEINERNAKIPTFLYAMPFSSNRIFLEETS 303
QY 203 YIDKANLQAEARQNIRDYAARQGWPLQTLRREQ-----GALPITLTGDNQFPWQQO 256
DB 304 LVAPGLKMEDIQERMAARLKHGLNVRKIEEDERCVPIMGGLPVL-----P 351
```

```

Y 257 QACSL--RAGLHPTTGYSLPLAVALADRLSALDVFSSVHQTIAH----- 302
b 352 ORVVGIGTAGMHPSTGY-----WVARTLAADIVANSIVQVLSVSDSLSGNDLSADV 405
Y 303 ----PAQRWQQGGRFRLNMLFLAGPAESRWVQRFYGLPEDLIARFYAGKLT 354
D 406 WDLWPIERRRQREF--CFGMILLKLDLGTGRRFFDAFF-----DLEPRYWHGFLS 456

RESULT 10
CYB TOBAC
D LCYB TOBAC STANDARD; PRT; 500 AA.
Q43578;
I 16-OCT-2001 (Rel. 40, Created)
I 16-OCT-2001 (Rel. 40, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
Lycopene beta cyclase, chloroplast precursor [EC 1.14.-.-].
LCY1 OR CRT1.1.
Nicotiana tabacum (Common tobacco).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
lamids; Solanales; Solanaceae; Nicotiana.
NCBI_TaxID=4097;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Samsun NN; TISSUE=Leaf;
MEDLINE=96434545; PubMed=8837512;
Cunningham F.A. Jr., Pogson B., Sun Z., McDonald K.A., Dellapenna D.,
Gantt E.;
"Functional analysis of the beta and epsilon lycopene cyclase enzymes
of Arabidopsis reveals a mechanism for control of cyclic carotenoid
formation";
Plant Cell 8:1613-1626(1996).
-!- FUNCTION: Catalyzes the double cyclization reaction which converts
lycopene to beta-carotene and neurosporene to beta-zeacarotene.
-!- PATHWAY: Carotenoid biosynthesis.
-!- SUBCELLULAR LOCATION: Chloroplast.
-!- SIMILARITY: Belongs to the lycopene cyclase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
EMBL; X81787; CAA57386.1; -.
PIR; S72506; S72506.
InterPro; IPR008671; Lycopene_cycl.
InterPro; IPR00205; NAD BS.
Pfam; PF05834; Lycopene_cycl; 1.
Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
Transit peptide.
TRANSIT 1 81 CHLOROPLAST (POTENTIAL).
CHAIN 82 500 LYCOPENE BETA CYCLASE.
NP_BIND 86 114 NAD (POTENTIAL).
SEQUENCE 500 AA; 56067 MW; 2E3721B87EB6CBC8 CRC64;

Query Match
Best Local Similarity 6.9%; Score 140; DB 1; Length 500;
Matches 83; Conservative 72; Mismatches 152; Indels 104; Gaps 20;

6 DLIVVAGLANGLIALRQQQHPDRILLIEAGPS---AGNHTW---SFHEED-LTLNQH 59
85 DLAVVGGGPAG--LAVAQVSEAGLSVSDPSKLIWPNNGVWVDFEAMDLDCDLA 142
60 RWIAPLVVHPDPVQRRPQRRHNSGYCVTSRHFAGILRQQFGHLW-----L 110
143 TWSGTGVV-----DIDNTTKDLDRPYGRVNRKQLSKMKQKILNGVXF 186
111 HTA--VSAVHAESVQLA---DGRITIHASTVIDRGYTP-----DSALRVGFQAFIQGEWQ 160

```

```

Db 187 HEAKVTKVTHEEAKSLKICNDGVTTQATVLDATGFSRCLVQYDKPFGYQVAYGILAE 246
Qy 161 LSA-PHGLSSPIIMD-----ATVDOQNGY--RFVYTLPLSATALLIBDTHYIDKAN 208
Db 247 VZEHPDTSKVLMDWRDHLGNMELAKERNKVTFLYAMPFSSNKIFLETSIVARPG 306
Qy 209 LQARARQNRIRYAARQGWPLQTLRBEQ-----GALPITLTGDNRFQWQQOQAC--S 260
Db 307 LRMDDIQERWVARLNLGIKVKYSIEDEHCVIPMGGSLEVI-----PQRVVGT 354
Qy 261 GLRAGLFHPTTCYSLSPLAVALADRLSALDVFSSSVH-----QTLAHFAQQRW- 308
Db 355 GGTAGLVHPSTGY-----MVARTLAAPVAVANILHYLGSEKOLLGNELSAWVKDLWP 408
Qy 309 ----QQQGFRLM-LNRMLFLAGPAESRWVQRFYGLPEDLIARFYAGKLT 354
Db 409 IERRRQREFCFGMILLKLDLPA-----TERRPDAFEDLEPRYWHGFLS 453

RESULT 11
LCYB CAPAN
ID LCYB CAPAN STANDARD; PRT; 498 AA.
AC Q43415;
Df 16-OCT-2001 (Rel. 40, Created)
Df 16-OCT-2001 (Rel. 40, Last sequence update)
Df 28-FEB-2003 (Rel. 41, Last annotation update)
Lycopene beta cyclase, chloroplast precursor [EC 1.14.-.-].
LCY1 OR CRT1.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Lamuyo; TISSUE=Fruit;
MEDLINE=96045549; PubMed=7550379;
Hugueney P., Badillo A., Chen H.C., Klein A., Hirschberg J.,
Camara B., Kuntz M.;
"Metabolism of cyclic carotenoids: a model for the alteration of this
biosynthetic pathway in Capsicum annuum chromoplasts.";
Plant J. 8:417-424(1995).
CC -!- FUNCTION: Catalyzes the double cyclization reaction which converts
lycopene to beta-carotene and neurosporene to beta-zeacarotene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- SIMILARITY: Belongs to the lycopene cyclase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X86221; CAA60119.1; -.
InterPro; IPR001327; FAD pyr redox.
InterPro; IPR008671; Lycopene_cycl.
InterPro; IPR00205; NAD BS.
InterPro; IPR001100; Pyr_redox.
Pfam; PF05834; Lycopene_cycl; 1.
PRINTS; PR00368; FADPNR.
PRINTS; PR00411; PNDRDTASEI.
Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast;
Transit peptide.
TRANSIT 1 79 CHLOROPLAST (POTENTIAL).
CHAIN 80 498 LYCOPENE BETA CYCLASE.
NP_BIND 84 112 NAD (POTENTIAL).
SEQUENCE 498 AA; 55610 MW; 177180CD5745F64F CRC64;

Query Match
Score 139; DB 1; Length 498;

```


"Gene structure and regulation of the carotenoid biosynthesis pathway in *Arabidopsis thaliana*.";
Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

[3] SEQUENCE FROM N.A.
STRAINE=cv. Columbia;
MEDLINE=20181125; PubMed=10718197;
Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Kotani H.,
Tabata S.,
"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
features of the regions of 3,076,755 bp covered by sixty pl and YAC
clones.",
DNA Res. 7:31-63(2000).

[4] SEQUENCE FROM N.A.
STRAIN=cv, Columbia;
MEDLINE=22954850; PubMed=14593172;
Yamada K., Iim J., Dalle J.M., Chen H., Shinn P., Palm C.J.,
Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
Char M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
Akazawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
Chao Q., Choy N., Enju A., Goldsmith A.D., Gurral M., Hansen N.F.,
Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
Xian S., Koeseena E., Ishida J., Jiang P.X., Jones T., Kawai J.,
Kaniya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.:
"Empirical analysis of transcriptional activity in the Arabidopsis
genome." Science 302:842-845(2003)

- !- FUNCTION: CATALYZES THE SINGLE CYCLIZATION REACTION WHICH CONVERTS LYCOPENE TO DELTA-CAROTENE AND NEUROSPORENE TO ALPHA-ZEACAROTENE.
- !- PATHWAY: Carotenoid biosynthesis.
- !- SUBCELLULAR LOCATION: Chloroplast.

-!- SMIILARITY: Belongs to the lycopene cyclase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcements> or send an email to license@isb-sib.ch).

ENBL; U50738; AAB53336.1; -;					Oxidoreductase; NAD; Carotenoid biosynthesis; Chloroplast; Transit peptide.
ENBL; AF117257; BAA82389.1; -;	TRANSIT	1	45	CHLOROPLAST (POTENTIAL).	
ENBL; AB024035; BA97033.1; -;	CHAIN	46	524	LYCOPENE EPSILON CYCLASE.	
ENBL; AY040024; AAK64181.1; -;	NP BIND	111	139	NAD (POTENTIAL).	
ENBL; AY079371; AAL85102.1; -;	CONFLICT	111	111	L -> H (IN REF. 1).	
InterPro; IPR001327; FAD pyr_redox.	SEQUENCE	524 AA;	58491 MW;	4CIF98CC72EDD074 CRC64;	
InterPro; IPR008671; Lycopene cycl.					
Pfam; PF05834; Lycopene cycl.; 1.					
PRINTS; PR00368; FADPNR-					

```

Query Match      6.4%; Score 129.5; DB 1; Length 524;
Best Local Similarity 20.0%; Pred: No. 0.0037;
Matches 81; Conservative 65; Mismatches 176; Indels 83; Gaps
6 DLIVGAGLANGLIARLQQQHPDMRILLI EAGPE---ACGNHTWSPHEEDLTIN---QH
110 DLVVIQGPAG--LALAESAKLGLKVGLI--GPDLPNTNNGYVWEDEFNDLGLQKCI EH
60 RWIAPLVVHHWPDYQV-----RPFQRHHVNSGYCVTSEHPAGILRQCFQHLWLHT:

```

166	VM-RETI	VLD	DDK	PIT	IGRAY	SVSR	RL	HEE	LL	RC	VE	---	GV	-----	SYL	SS	212	
113	AVS	AV	H	ASS	-----	V	QL	AD	GR	I	I	E	A	S	T	V	150	
213	KVDS	I	T	E	A	S	D	G	L	R	V	A	C	D	N	N	V	272
161	L-SAPH	GL	S	S	P	I	I	M	A	T	V	D	Q	N	G	Y	R	203
273	VENS	P	V	D	Q	M	V	E	N	D	-----	---	Y	D	T	N	E	324
204	ID	K	N	L	Q	A	E	R	A	S	O	N	I	D	Y	A	A	263
325	A	S	K	D	V	M	P	D	L	L	T	K	M	L	R	D	T	380
264	AGL	F	I	P	T	T	G	S	L	A	P	A	L	A	D	R	S	312
381	AS	M	V	E	P	A	T	G	S	V	R	S	L	E	A	P	Y	440
313	F	R	M	L	N	R	L	F	L	A	G	A	E	S	R	W	Q	357
441	P	F	-	L	G	L	A	L	I	V	O	P	E	T	E	G	I	484

RESULT 14
FIXC_ECOLI STANDARD; PRT; 428 AA.
ID FIXC_ECOLI
AC P31575; P75626;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE FIXC protein.
GN FIXC OR B0043 OR C0053.
OS Escherichia coli, and
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562, 217992;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region";
RL Nucleic Acids Res. 20:3305-3308(1992).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=23288234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Zhou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
[4]
RN SEQUENCE OF 1-188 FROM N.A.
RC STRAIN=O44:K74;
RX MEDLINE=96066354; PubMed=7473063;
RA Bichler K., Buchet A., Bourgis F., Kleber H.-P.,
RA Mandrand-Berthelot M.-A.;

"The fix Escherichia coli region contains four genes related to

carnitine metabolism".

J. Basic Microbiol 35:217-227(1995).

- FUNCTION: PROBABLY ACCEPTS ELECTRONS FROM FIXA/FIXE AND REDUCES A

QUINONE.

- COFACTOR: FAD (potential).

- SIMILARITY: Belongs to the BTP-QO / fixC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D10483; BAB96611.1; -
EMBL; AB000114; AAC73154.1; -
EMBL; AB016755; AAN78549.1; -
PIR; X71977; CAA50799.1; -
PIR; C64725; C64725.
EcoGene; EGI1564; fixC.
InterPro; IPR001327; FAD pyr_redox.
InterPro; IPR000205; NAD_B5.
InterPro; IPR001100; Pyr_redox.
InterPro; IPR00103; Pyridine redox 2.
PRINTS; PR00368; FADPNR.
PRINTS; IPR003042; Rng_mnoxygenase.
PRINTS; PR00411; PNDROTASII.
PRINTS; PR00469; PNDROTASII.
PRINTS; PR00420; RNMNOXGNASE.
Oxidoreductase; FAD; Flavoprotein; Complete proteome.
NP_BIND 12 28 FAD OR NAD(P) (ADP PART) (POTENTIAL).
CONFLICT 1 3 MSE -> MT (IN REF. 4).
CONFLICT 64 64 D -> E (IN REF. 1).
CONFLICT 280 281 VW -> IM (IN REF. 1).
CONFLICT 310 310 L -> F (IN REF. 1).
SEQUENCE 428 AA; 45702 MW; 4D089BESA30C7FBC CRC64;

Query Match 5.6%; Score 113.5; DB 1; Length 428;

Best Local Similarity 20.0%; Pred. No. 0.061;

Matches 88; Conservative 56; Mismatches 167; Indels 129; Gaps 17;

5 YDLIVGAGLGLIALRLQOHPDMRILLIAGPAGGNGHTWSFHEDLTLNQRWIAP 64
6 FDLIVGAGLGLSVAALVLARE--GAQLVIERGNSAGAKN-----VTGSELVAH 53
65 LVVHWPDYQVRPPQRR--RHVNSGY-----YC-----VTSRHFAGILROOFGOH 107
54 SLEHIIICFADSAFVERLITHEKLAFTWKSAMTNDYCNQDSTSPQRSYSVLRSKF--D 111
108 LMLHTAVSAVHAESVQ-----LADGRIIHASTVIDGRGYPDSALRVG 150
112 AMLMQAEAGAGLITGIRVDNIVORDGKVGVEADGVIEAKTVILADGVNSILAEKIG 171
151 F-----QAFIQEWOLSAFGLSSPIIMDATVDOQNGYRFFVYILP 190
172 MAKRVKPTDVAVGKELIELPKSVIEDRFQLOQNOG--AACLFAGSPDCLDWGGFLYINE 230
191 LSATALLIEITHYIDKANLQAEARQNIQDYARQGWPL---QTLIREQGLP-----241
231 NTLISGLVLCGLHLDKAKSVQPMLEDFKQHPAVA--PLIAGKLVESAHVVPPEAGINM 288
242 -----ITLTGNRQFWQQQQAQSGLAAGLFHTTGYSLPLVALADRLSALDVFTS 293
289 LPELVGQVLIAGD-----AAGCMNIGFTIRGMDLAIAAGEAAKTVLSMKVS 337
294 SSVHQTAAHQQRWQQGQFFRMNMLFLAGAESRWRMQRFVGLPDLILARFYAGKL 353
338 DD-----FSKQLAEY-----RQHLESGPL-----EDMRMYQLPAFL-----370
354 TVTDRLRLSLGKPPVPVFAA 373

Db 371 --DNPRMFGYPPELAVGVA 387

RESULT 15

Y801 DEIRA

ID Y801 DEIRA STANDARD; PRT; 410 AA.

AC Q9RW68;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative carotenoid cyclase DR0801.

GN DR0801.

OS Deinococcus radiodurans.

OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

CX NCBI_TaxID=1299;

IL

RN SEQUENCE FROM N.A.

RP STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;

RX MEDLINE=20036896; PubMed=10567286;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

radiodurans R1.";

RL Science 286:1571-1577(1999).

CC - SIMILARITY: Belongs to the lycopene cyclase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB001934; AAF10377.1; -

PIR; D75475; D75475.

TIGR; DR0801; -

DR InterPro; IPR008671; Lycopene cycl.

DR InterPro; IPR003042; Rng_mnoxygenase.

DR Pfam; PF05834; Lycopene cycl. 1.

DR PRINTS; PR00420; RNMNOXGNASE.

KW Hypothetical protein; Oxidoreductase; NAD; Carotenoid biosynthesis;

KW Complete proteome.

FT NP_BIND 11 39 NAD (POTENTIAL).

SQ SEQUENCE 410 AA; 43161 MW; E1B3162F10F9E6AF CRC64;

Query Match 5.3%; Score 106.5; DB 1; Length 410;

Best Local Similarity 22.8%; Pred. No. 0.22; Indels 121; Gaps 22;

Matches 100; Conservative 44; Mismatches 173;

6 DLIIVGAG-----LANGLIARLQOQ-----HPDMRILLIENGPEAGNGHTWSFHEDLTL 56

10 DVLVIGGPGSGTALSAAELAAAGLDVQLAPHP-RPPATYCAWLGLDPTWA-----60

57 NOHRWIAPIVHHWPDYQVRPPQRRHVSNGYCVTSRHFAGILROOFGHLLWHTAVSA 116

61 ---RGCAEQV---WTDVRAVYTGPTSLGQFVALLDN---AALLRTLRLGLADMTWVEGHA 111

117 VHAEVSQIADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQL-----SAPHG-LSSP 170

112 LHAER-----SGAGWTIVYGA-----GGERWQTVLVDASGHALVSP 148

171 IIMDATVDOQNGY-----RFTYTLPLSATALL 197

149 VRFFGGAAQLQATYGVVARFRPVPVTPGSMVMMDYRTPAPELKEGATFLYAMHLGDRYF 208

198 IEDTHYIDK-ANLQAEARQNIQDYARQGWPLQTLIREQCALPITLTGDNRFQWQOOP 256

209 VEETSLIARPAAPTRAELELRLRLISA-QCTPPHATESSEWVAFPM-----NAQAPAPGG 262
257 QACSLRAGLFHPTTGYSLPLAVLADRLSALDVTSSSVHQTIAHFA--QORWQQQGF 314
263 VLAYGAAGRVHPVSGFQ--VAGALSDFGVATATATACQGGKDAAGWAALWSPE--- 317
315 RMLNRLFLAG-----PAESRWVMQRFYGLPEDLIARFY-----AGKLTVTDLRI 361
318 RRAAREVHLIGVALLGLERAEPL-HFFGTFFGLPREQWARFIHPDPTDAGTLART-MLRV 375
362 LS---GKPPVVFPAALQA 376
376 FAQTGGVRVLPFLARAALA 393

Search completed: February 29, 2004, 14:45:10
Job time : 10.4382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

W protein - protein search, using sw model

Run on: February 29, 2004, 14:34:14 ; Search time 10.6328 Seconds
(without alignments)
3455.835 Million cell updates/sec

Title: US-09-941-947A-30

Perfect score: 2021
Sequence: 1 MQPHYDLILVAGLANGLIA.....SGKPPVPVFAALQAIMTTHR 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1715	84.9	382	2 S52585	dycopene cyclase -
2	1678	83.0	382	2 C37802	crty protein - Erw
3	1148.5	56.8	386	2 S52981	lycopene cyclase -
4	152	7.5	526	2 T07082	lycopene epsilon-c
5	148.5	7.3	498	2 S71511	capsanthin/capsoru
6	147.5	7.3	471	2 S51511	capsanthin/capsoru
7	140	6.9	500	2 S72506	lycopene beta-cycl
8	139	6.9	500	2 S66349	lycopene beta-cycl
9	133	6.6	500	2 S72505	lycopene beta-cycl
10	133	6.6	500	2 S66350	lycopene beta-cycl
11	127.5	6.3	399	1 E70888	probable glf prote
12	117.5	5.8	391	2 AD0584	probable monooxyge
13	114.5	5.7	428	2 F06334	flavoprotein [impo
14	114.5	5.7	428	2 F85485	flavoprotein, elec
15	114.5	5.7	504	2 G87532	cryptophan halogen
16	113.5	5.6	428	2 C64725	fixC protein - Esc
17	110.5	5.5	405	2 T37022	probable lycopene
18	108.5	5.4	3161	2 T30342	protein HWP1 - Ye
19	106.5	5.3	410	2 D75475	lycopene cyclase -
20	105	5.2	398	2 D83122	probable FAD-depen
21	105	5.2	557	2 F83181	probable GMC-type
22	104.5	5.2	402	2 AC2797	2-octaprenyl-6-met
23	104.5	5.2	402	2 C97576	2-octaprenyl-6-met
24	104	5.1	413	2 D86920	probable UDP-galac
25	103.5	5.1	722	2 B75608	GMC oxidoreductase
26	103.5	5.1	3163	2 T17440	probable polyketid
27	103.5	5.1	3163	2 AG0233	yersiniabactin bio
28	102.5	5.1	392	2 AC0873	2-octaprenyl-6-met
29	101.5	5.0	391	1 D64801	probable monooxyge

ALIGNMENTS

RESULT 1

S52585

dycopene cyclase - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 10-Dec-1999

C:Accession: S52585

R:Lin, Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52585

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-382 <LIN>

A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21262.1; PID:G148396

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992

C:Superfamily: Erwinia dycopene cyclase

Query Match 84.9%; Score 1715; DB 2; Length 382;

Best Local Similarity 83.8%; Pred. No. 1.6e-135;

Matches 320; Conservative 31; Mismatches 31; Indels 0; Gaps 0;

Qy	1	MQPHYDLILVAGLANGLIALRQQHPDMRILLIEAGPEAGGNHTWSPHEEDLTINQHR	60
Db	1	MRPHYDLILVAGLANGLIALRQQHPDMRILLIEAAPQAGCNHTWSPHDALTSSQHR	60
Qy	61	WIAPLVVHWHPDYQVRFPPQRRHVSQYCVTSRHPAGILRQFGQHLWLTAVSAVHAE	120
Db	61	WVAPLVVHWHPDYQVRFPPQRRKLSNGYFVTSQRFABVLQQFQGHLMISRAVAEVHAD	120
Qy	121	SVQLADGRITHASTVIDGRGTPDPSALRVGFQAFIQEWQLSAPHGLSSPIMDATVQQ	180
Db	121	AVLANNQVISAVIDGRGTPNSALNVGFQAFIQEWFLSKPHGLSSPIMDATVQQ	180
Qy	181	NGYRFVYIPLSATALLIEDTHYIDKANLQARRAQNRDVAARQGWPLQTLRBEQAL	240
Db	181	NGYRFVYISLFLSATALLIEDTHYIDNATLEPERAQNRDVAARQGWQWLQTLRBEQAL	240
Qy	241	PITLTGDNQFQQQPOQACGLRAGLFFPTTGYSLAVALADRLSALDVFTSSSVHTI	300
Db	241	PITLTGDSVAFWQQQPLACGLRAGLFFPTTGYSLAVALADRLSALDVFTSSSHQAI	300
Qy	301	AHPAQRRWQQQGFPMNLNMLFLAGPAESRRVRWQRFYGLPEDLIARFYAGKLTVDRLR	360
Db	301	THFAHERWQQQGFPMNLNMLFLAGPADSRVRWQRFYGLPEDLISRFYAGKLTVDRLR	360
Qy	361	ILSGKPPVPVFAALQAIMTTHR	382
Db	361	ILSGKPPVPVLAALQAIMTTHR	382

RESULT 2

C37802
 cr1y protein - Erwinia uredovora
 C:Species: Erwinia uredovora
 C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
 C:Accession: C37802
 R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashina, J.; Bacteriol. 172, 6704-6712, 1990
 J: Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional genomics
 A:Reference number: A37802; MUID:91072214; PMID:2254247
 A:Accession: C37802
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-382 <MIS>
 A:Cross-references: GB:D90087; NID:g216681; PIDN:BAAL4126.1; PID:g216684
 C:Superfamily: Erwinia dycopene cyclase

Query Match 83.0%; Score 1678; DB 2; Length 382;
 Best Local Similarity 82.2%; Pred. No. 1.9e-132;
 Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

2Y 1 MOPHYDLILVGGGLANGLIARLQYRYPOLNLLIEAGQPCCGNTWTFHEDDLTPGOHAWLAPL 60
 DB 1 MOPHYDLILVGGGLANGLIARLQYRYPOLNLLIEAGQPCCGNTWTFHEDDLTPGOHAWLAPL 60
 2Y 61 MIAPLVVHHWPDYQVRRPQRRHNSGYCYVTSRHFAGILRQFQGOHLWHTAVSAVHAE 120
 DB 61 MIAPLVVHHWPDYQVRRPQRRHNSGYCYVTSRHFAGILRQFQGOHLWHTAVSAVHAE 120
 2Y 121 SVQLADGRIIHASTVIDRGYTPDSALRVGFCAGIAGWQSLAPHGLSSPIINDATVDQ 180
 DB 121 SVQLADGRIIHASTVIDRGYTPDSALRVGFCAGIAGWQSLAPHGLSSPIINDATVDQ 180
 2Y 181 NGYRFPVYSLPTELLIEDTHYIDKAMQLAERARQNIYAAQGWPLQTLREEQGAL 240
 DB 181 NGYRFPVYSLPTELLIEDTHYIDKAMQLAERARQNIYAAQGWPLQTLREEQGAL 240
 2Y 241 PITLGDNRQWQOQPCQGLRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTIAHFA 300
 DB 241 PITLGDNRQWQOQPCQGLRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTIAHFA 300
 2Y 301 AHFAQRRQOQOQFFFMNLNMLFLAGPAESRWVMOVFYGLPDLIARFYAGKLTVTDLRL 360
 DB 301 THFAERWQOQOQFFFMNLNMLFLAGPADSRWVMOVFYGLPDLIARFYAGKLTVTDLRL 360
 2Y 361 ILSGKPPVPVPAALQAIMTTHR 382
 DB 361 ILSGKPPVPVPAALQAIMTTHR 382

RESULT 3
 S22981
 lycopen cyclase - Erwinia herbicola
 C:Species: Erwinia herbicola
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
 C:Accession: S52981
 R: Hundle, B.; Alberti, M.; Niveltstein, V.; Beyer, P.; Kleinig, H.; Armstrong, G.A.; Burkolter, Genet. 245, 406-416, 1994
 J: Title: Functional assignment of Erwinia herbicola Eho 10 carotenoid genes expressed in Agrobacterium
 A:Reference number: S52976; MUID:95107236; PMID:7808389
 A:Accession: S52981
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-386 <HUN>
 A:Cross-references: ENBL:M87280; NID:g148404; PIDN:AAA64980.1; PID:g148411
 C:Genetics:
 C:Start codon: GTG
 C:Superfamily: Erwinia dycopene cyclase

Query Match 56.8%; Score 1148.5; DB 2; Length 386;
 Best Local Similarity 58.3%; Pred. No. 3.3e-88;
 Matches 218; Conservative 53; Mismatches 102; Indels 1; Gaps 1;

2Y 6 DLIIVGGLANGLIARLQYRYPOLNLLIEAGQPCCGNTWTFHEDDLTPGOHAWLAPL 65

DB 3 DLIIVGGLANGLIARLQYRYPOLNLLIEAGQPCCGNTWTFHEDDLTPGOHAWLAPL 62
 QY 66 VVHHWPDYQVRRPQRRHNSGYCYVTSRHFAGILRQFQGOHLWHTAVSAVHAEVOL 125
 DB 63 VAHAMPGYEVQPPDLRLRLAGYTSISERFAEALHQAALGENIWLNCSEVLPNSVRLA 122
 QY 126 DGRIIHASTVIDRGYTPDSALRVGFCAGIAGWQSLAPHGLSSPIINDATVDQNGYRF 185
 DB 123 NGEALLAGAVIDRGYTPASSAMQTYQLFQOQWRLTQPHGLTVPILMDATVAQOQGYRF 182
 QY 186 VYTLPLSATALLIEDTHYIDKAMQLAERARQNIYAAQGWPLQTLREEQGALPITLT 245
 DB 183 VYTLPLSATALLIEDTHYIDKAMQLAERARQNIYAAQGWPLQTLREEQGALPITLT 242
 QY 246 GDNRFQWQOQPCQGLRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTIAHFA 304
 DB 243 GDIQALWADAPGVPRSGMRAGLPHPTTGYSLPLAVALADRLSALDVTSSVHTIAHFA 302
 QY 305 QORWQOQOQOQFFFMNLNMLFLAGPAESRWVMOVFYGLPDLIARFYAGKLTVTDLRL 364
 DB 303 EXHMRQOQOQFFFMNLNMLFLAGPAESRWVMOVFYGLPDLIARFYAGKLTVTDLRL 362
 QY 365 KPPVPVPAALQAIM 378
 DB 363 KPPVPVPAALQAIM 376

RESULT 4
 T07082
 lycopen epsilon-cyclase (EC 5.5.1.1) - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2003
 C:Accession: T07082
 R: Roreen, G.; Cohen, M.; Zamir, D.; Hirschberg, J.
 submitted to the EMBL Data Library, July 1997
 A:Description: Regulation of expression of the gene for lycopen epsilon cyclase during
 A:Reference number: Z15905
 A:Accession: T07082
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-526 <RON>
 A:Cross-references: ENBL:Y14387; NID:el285211; PIDN:CAA74745.1; PID:el285212
 A:Experimental source: cultivar VF36; leaf
 C:Genetics:
 A:Gene: Crtl-e-1
 A:Map position: 12
 C:Function:
 A:Description: converts lycopen to delta-carotene
 C:Superfamily: tomato lycopen beta-cyclase
 C:Keywords: intramolecular lyase; isomerase

Query Match 7.5%; Score 152; DB 2; Length 526;
 Best Local Similarity 21.1%; Pred. No. 8.9e-05;
 Matches 87; Conservative 66; Mismatches 167; Indels 92; Gaps 17;

QY 6 DLIIVGGLANGLIARLQYRYPOLNLLIEAGQPCCGNTWTFHEDDLTPGOHAWLAPL 62
 DB 107 DLWVGCGPAG--LALAAEAKUGLANVGLV--GPDLPFTNNYGVWDEDFKDLGLQ----- 157
 QY 63 APLVHHWPDYQVRRPQRRHNSGYCYVTSRHF-----AGILRQFQGOHLW 111
 DB 158 -ACIEHWRTDITVYLDDBDDEPILIGRAYGRVSRHFLHEILKRCVEAGV-----LYLN 208
 QY 112 T-----AVSAVHAEVOLADGRIL---HASTVIDRG-----YTPDSALRVGFAPTGOE 158
 DB 209 SKVDRIEATNGQSLVECEGDWIPCFVTVASGAASGKFLQVLELQSP-RVSVQTAIGVE 267
 QY 159 WQL-SAPHGLSSPIINDATVDQNGYR-----FVYTLPLSATALLIEDT 201
 DB 268 VEVDNNPFDLSLVFMD-----YDYLHDAQSLAEAKYPTFLYAMFMSPTRVPEET 319
 QY 202 HYIDKAMQLAERARQNIYAAQGWPLQTLREEQGALPITLTGDNRFQWQOQPCQSG 261

320 CLASDAMPFLLKKMLRLNTLGLRIKEIYEWWSYIPGGSLPA-----TQKTLAFG 375
262 LRAGLFHTTGYSLPLAVALADRLSAL-----DVFTSSVHEQTIAH-----FAQ 305
376 AAASVMVHPATGYSVVRSLEAPKCAVLANTLROHYSKNMLTSSSIPISTQAMNTLWPQ 435
306 QRWQCGQFFRMLNRLFLAGPABSRWRVMQRFYGLPDLIARFVAGKLTWTD 357
436 ERKQRSEF-LFGLALILQLDIEGRSFFRAFFRVPKMMWQGLGSSLSAD 486

RESULT 5
1511
capsanthin/capsorubin synthase (EC 5.5.1.-) - pepper
Alternate names: chromoplast-specific protein
Contains: capsorubin synthase; capsanthin synthase
Species: Capsicum annuum (pepper)
Date: 09-Dec-1997 #sequence_revision 09-Dec-1997 #text_change 28-Jul-2000
Accession: S71511; J02141
Bouvier, F.; Hugueney, P.; d'Harlingue, A.; Kuntz, M.; Camara, B.
ant J. 6, 45-54, 1994
Title: Xanthophyll biosynthesis in chromoplasts: isolation and molecular cloning of an
Reference number: S71511; MUID:95004653; PMID:7920703
Accession: S71511
Molecule type: mRNA
Residues: 1-498 <BD>
Cross-references: EMBL:X76165; NID:9522119; PIDN:CAA53759.1; PID:9522120
Experimental source: cv. Lamuyo; developmental stage ripening fruits
Deruere, J.; Bouvier, F.; Steppuhn, J.; Klein, A.; Camara, B.; Kuntz, M.
ochem. Biophys. Res. Commun. 199, 1144-1150, 1994
Title: Structure and expression of two plant genes encoding chromoplast-specific prote
Reference number: J02140; MUID:94197697; PMID:8147854
Accession: J02141
Molecule type: DNA
Residues: 1-498 <DB>
Cross-references: EMBL:X77289; NID:9468747; PIDN:CAA54495.1; PID:9468748
Experimental source: cv. Yolo Wonder
Deruere, J.; Bouvier, F.; Steppuhn, J.; Klein, A.; Camara, B.; Kuntz, M.
ochem. Biophys. Res. Commun. 201, 486, 1994
Reference number: A55775
Contents: annotation; erratum
Note: the legends of the nucleotide sequence figures were reversed in publication (see
Genetics:
Gene: CCS
Complex: monomer
Function:
Description: catalyses the conversion of the ubiquitous 5,6-epoxycartenoids, antheraxa
Pathway: carotenoid biosynthesis
Superfamily: tomato lycopene beta-cyclase
Keywords: chromoplast; intramolecular lyase; isomerase

Query Match 7.3%; Score 148.5; DB 2; Length 498;
Best Local Similarity 22.2%; Pred. No. 0.00016;
Matches 72; Conservative 58; Mismatches 125; Indels 69; Gaps 16;
5 YDLILVAGLGLIARLQOQ---HPDMRILLIAGPEA---GQNTWTFHEEDLTN-- 57
32 FDVLIIGTGA---GLRLAEQVSKYGIKVCVDPSPSLMWPNNYGVWVDFEKLGLDC 137
58 -QHWIAPLVVHPDPYQVFPQ---RHVNSGYCVTSR---HFAGILRQ 102
138 LDHKK---PVSCVHLSHDKTKYLDPRYGRVSRKLLKLLNS---CVENRVKPYKAKVLKV 192
103 QFGOHLMLHTAVSAVHAB---SVOLADGRIIHAFTVIDGRGYTP-----DSALRVGFOAF 154
193 K-----HBEFESSIVCDDGRKISGLSIVDASGYASDFIEYDKPRNHGYQVA 238
155 IGQEWQL-SAPHGLSSPIIND-----ATVDOQNGYRFVYTLPLSATALLIEDTH 202
239 HGILAEVDNHPFDLQKMLMDWRDHLGNBPYLRVKNKTEPTFLYAMPFDRLNVLFLBETS 298
203 YIDKANLQAEARQNIADYARQGWPLQTLIREEQGALPITLTGDNRFQWQOQFQACSG 262
299 LVSRPMLSYKVEKRWYRLRHILGILKYSVLEEK--CVITMGGLPRI-PQNVMAIGG- 354
263 RAGLFHTTGYSLPLAVALA 282
355 TSGIVHPSSGYMVARSMALA 374

RESULT 7
S72506
lycopene beta-cyclase (EC 5.5.1.-) - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 28-Jul-2000
Accession: S72506
R:Pecker, I.; Gabbay, R.; Cunningham Jr., P. X.; Hirschberg, J.
submitted to the EMBL Data Library, September 1994
A:Description: Cloning and characterization of the lycopene cyclase gene from tomato.
A:Reference number: S72506
A:Status: preliminary
A:Accession: S72506
A:Molecule type: DNA
A:Residues: 1-500 <PBC>
A:Cross-references: EMBL:X81787; NID:g1006689; PID:g1006690
C:Superfamily: tomato lycopene beta-cyclase
C:Keywords: intramolecular lyase; isomerase

299 LVSRPMLSYKVEKRWYRLRHILGILKYSVLEEK--CVITMGGLPRI-PQNVMAIGG- 354
263 RAGLFHTTGYSLPLAVALADRLS 286
355 TSGIVHPSSGYMVARSMALAPVLA 378

RESULT 6
S51511
capsanthin/capsorubin synthase (EC 5.5.1.-) - pepper
C:Species: Capsicum annuum (pepper)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 28-Jul-2000
Accession: S51511
R:Houline, G.; Schantz, M. L.; Meyer, B.; Pozueta-Romero, J.; Schantz, R.
Curr. Genet. 26, 524-527, 1994
A:Title: A chromoplast-specific protein in Capsicum annuum: characterization and express
A:Reference number: S51511; MUID:95179811; PMID:787474
Accession: S51511
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-471 <BD>
A:Cross-references: GB:X78030; NID:9840728; PIDN:CAA54961.1; PID:9840729
C:Superfamily: tomato lycopene beta-cyclase
C:Keywords: intramolecular lyase; isomerase

Query Match 7.3%; Score 147.5; DB 2; Length 471;
Best Local Similarity 22.2%; Pred. No. 0.00018;
Matches 71; Conservative 57; Mismatches 123; Indels 69; Gaps 16;
5 YDLILVAGLGLIARLQOQ---HPDMRILLIAGPEA---GQNTWTFHEEDLTN-- 57
82 FDVLIIGTGA---GLRLAEQVSKYGIKVCVDPSPSLMWPNNYGVWVDFEKLGLDC 137
58 -QHWIAPLVVHPDPYQVFPQ---RHVNSGYCVTSR---HFAGILRQ 102
138 LDHKK---PVSCVHLSHDKTKYLDPRYGRVSRKLLKLLNS---CVENRVKPYKAKVLKV 192
103 QFGOHLMLHTAVSAVHAB---SVOLADGRIIHAFTVIDGRGYTP-----DSALRVGFOAF 154
193 K-----HBEFESSIVCDDGRKISGLSIVDASGYASDFIEYDKPRNHGYQVA 238
155 IGQEWQL-SAPHGLSSPIIND-----ATVDOQNGYRFVYTLPLSATALLIEDTH 202
239 HGILAEVDNHPFDLQKMLMDWRDHLGNBPYLRVKNKTEPTFLYAMPFDRLNVLFLBETS 298
203 YIDKANLQAEARQNIADYARQGWPLQTLIREEQGALPITLTGDNRFQWQOQFQACSG 262
299 LVSRPMLSYKVEKRWYRLRHILGILKYSVLEEK--CVITMGGLPRI-PQNVMAIGG- 354
263 RAGLFHTTGYSLPLAVALA 282
355 TSGIVHPSSGYMVARSMALA 374

Query Match 6.9%; Score 140; DB 2; Length 500;
Best Local Similarity 20.2%; Pred. No. 0.0084;
Matches 83; Conservative 72; Mismatches 152; Indels 104; Gaps 20;
Y 6 DLILVAGLANGIALRLOQHPDMRILLIEAGPE---AGGNHTW---SPHEED-LTLNQH 59
b 85 DLAVVGGGPAG--LAVAQVSEAGLSVSDPSKLIWPNNGVWVWDEFEAMDLLDCLDA 142
Y 60 RTIAPLVVHWPYQVRFPPQRRHNSGYCVTSRHFAGILRQOFGQHLW-----L 110
b 143 TWSGVV-----YIDNTTKDLDRPYGRVNRKQLKSKMOKCILNGVKF 186
Y 111 HTA--VSAVHAESVOLA---DGRHASTVIDRGYTP-----DSALRVGFAFTQEWQ 160
b 187 HPAKVIHIEEAKSMILCNDGVITQATVLDATGSRCLVQYDPRKGYQVAVGILAE 246
Y 161 LSA-PHGLSSPIIMD-----ATVQONGY--RFVYTLPLSATALLIEDTHYIDKAN 208
b 247 VEEHPDTSKWLMDWEDSHLGNMELKERNKRVPTFLVAMPFSSNKIPLEETSLVAPG 306
Y 209 LOARARONRDYAAQGWPLQTLREEQ-----CALPITLTGDNRPQWQOQAC--S 260
b 307 LRWDDIQERWVARLNLHGIKVSIEDEHCVIPMGSLPVI-----PQVVGT 354
Y 261 GLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVH-----QTIAPFAQQRW- 308
b 355 GGAGLVHPSTGY-----MVAARTLAAAPVANAIIHYLGSEKLLGNLSAAVWKDLP 408
Y 309 ----QQQGFIRM-LNRMLFLAGPAESRWRVMOFYGLPEDLIARFYAGKLT 354
b 409 IERRQRBEFFCGMDILLKLDLPA-----TRFFDAFFDLPRYWHGFLS 453

RESULT 8
66349
ycopene beta-cyclase (EC 5.5.1.1) - common tobacco
;Species: Nicotiana tabacum (common tobacco)
;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 28-Jul-2000
;Accession: S66349
;Title: Cloning and characterization of the cDNA for lycopen beta-cyclase from tomato
;Reference number: S66349; MUID:96194462; PMID:8624411
;Accession: S66349
;Status: preliminary; nucleic acid sequence not shown
;Molecule type: DNA
;Residues: 1-500 <PEC>
;Cross-references: EMBL:X81787
;Superfamily: tomato lycopen beta-cyclase
;Keywords: intramolecular lyase; isomerase

Query Match 6.9%; Score 139; DB 2; Length 500;
Best Local Similarity 20.2%; Pred. No. 0.001;
Matches 83; Conservative 72; Mismatches 152; Indels 104; Gaps 20;
Y 6 DLILVAGLANGIALRLOQHPDMRILLIEAGPE---AGGNHTW---SPHEED-LTLNQH 59
b 85 DLAVVGGGPAG--LAVAQVSEAGLSVSDPSKLIWPNNGVWVWDEFEAMDLLDCLDA 142
Y 60 RTIAPLVVHWPYQVRFPPQRRHNSGYCVTSRHFAGILRQOFGQHLW-----L 110
b 143 TWSGVV-----YIDNTTKDLDRPYGRVNRKQLKSKMOKCILNGVKF 186
Y 111 HTA--VSAVHAESVOLA---DGRHASTVIDRGYTP-----DSALRVGFAFTQEWQ 160
b 187 HPAKVIHIEEAKSMILCNDGVITQATVLDATGSRCLVQYDPRKGYQVAVGILAE 246
Y 161 LSA-PHGLSSPIIMD-----ATVQONGY--RFVYTLPLSATALLIEDTHYIDKAN 208
b 247 VEEHPDTSKWLMDWEDSHLGNMELKERNKRVPTFLVAMPFSSNKIPLEETSLVAPG 306
Y 209 LOARARONRDYAAQGWPLQTLREEQ-----CALPITLTGDNRPQWQOQAC--S 260
b 307 LRWDDIQERWVARLNLHGIKVSIEDEHCVIPMGSLPVI-----PQVVGT 354
b 409 IERRQRBEFFCGMDILLKLDLPA-----TRFFDAFFDLPRYWHGFLS 453

Db 307 LRDDIQERWVARLNLHGIKVSIEDEHCVIPMGSLPVI-----PQVVGT 354
Qy 261 GLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVH-----QTIAPFAQQRW- 308
Db 355 GGAGLVHPSTGY-----MVAARTLAAAPVANAIIHYLGSEKLLGNLSAAVWKDLP 408
Qy 309 ----QQQGFIRM-LNRMLFLAGPAESRWRVMOFYGLPEDLIARFYAGKLT 354
Db 409 IERRQRBEFFCGMDILLKLDLPA-----TRFFDAFFDLPRYWHGFLS 453

RESULT 9
S72505
lycopene beta-cyclase (EC 5.5.1.1) - tomato
;Species: Lycopersicon esculentum (tomato)
;Date: 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 28-Jul-2000
;Accession: S72505
;Title: Cloning and characterization of the lycopene cyclase gene from tomato.
;Reference number: S72505
;Accession: S72505
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-500 <PEC>
;Cross-references: EMBL:X86452; NID:g1006672; PID:g1006673
;Superfamily: tomato lycopen beta-cyclase
;Keywords: intramolecular lyase; isomerase

Query Match 6.6%; Score 133; DB 2; Length 500;
Best Local Similarity 19.4%; Pred. No. 0.0032;
Matches 80; Conservative 74; Mismatches 152; Indels 106; Gaps 19;
Qy 6 DLILVAGLANGIALRLOQHPDMRILLIEAGPEA--GGNH----- 45
Db 85 DLAVVGGGPAG--LAVAQVSEAGLSVCSIDPNPKLIWPNNGVWVWDEFEAMDLLDCLDA 142
Qy 46 TWS-----FHEDITLQHRNIAPLVVHWPYQVRFPPQRRHNSGYCVTSRHFAGILR 101
Db 143 TWSGAAYVIDNTAKDLHR-----PYGRVNRKQLKSKMOK--KCMN----- 182
Qy 102 QQFGQHLWLTAVSAVHAESVOLA---DGRHASTVIDRGYTP-----DSALRVGFOA 153
Db 183 ---GVRFHQAKVIKVIHEESKSMILCNDGITQATVLDATGSRSLVQYDPRKPYNGYQV 239
Qy 154 PIQEWOLSA-PHGLSSPIIMD-----ATVQONGY--RFVYTLPLSATALLIEDT 201
Db 240 AYGILAEVVEHPDTSKWLMDWEDSHLGNMELKERNKRVPTFLVAMPFSSNRIPLEET 299
Qy 202 HYIDKANIQARARONRDYAAQGWPLQTLREEQ-----CALPITLTGDNRPQWQOQ 255
Db 300 SLVAPGLRIDDIQERWVARLNLHGIKVSIEDEHCVIPMGSLPVL----- 347
Qy 256 PQACSGI--RAGLFHPTTGYSLPLAVALADRL-SALDVFTSSSVHQTIAHFAQQRW---- 308
Db 348 PQRVVGGIGTAGVHPSTGYWVARTLAAAPVANAIIQYLGSEKSGNELSTAVWKOLW 407
Qy 309 ----QQQGFIRM-LNRMLFLAGPAESRWRVMOFYGLPEDLIARFYAGKLT 354
Db 408 PIERRQRBEFFCGMDILLKLDLPA-----TRFFDAFFDLPRYWHGFLS 453

RESULT 10
S66350
lycopene beta-cyclase (EC 5.5.1.1) - tomato
;Species: Lycopersicon esculentum (tomato)
;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 28-Jul-2000
;Accession: S66350
;Title: Cloning and characterization of the cDNA for lycopen beta-cyclase from tomato
;Reference number: S66349; MUID:96194462; PMID:8624411
;Accession: S66350

;Status: preliminary; nucleic acid sequence not shown.

;Molecule type: mRNA

;Residues: 1-500 <PEC>

;Cross-references: EMBL:X86452

;Superfamily: tomato lycopene beta-cyclase

;Keywords: intramolecular lyase; isomerase

Query Match 6.6%; Score 133; DB 2; Length 500;
Best Local Similarity 19.4%; Pred. No. 0.0032;
Matches 80; Conservative 74; Mismatches 152; Indels 106; Gaps 19;

Y 6 DLIVGAGLANGLIARLQQQHPDKXILLIEAGPEA--GGNH-----45

b 85 DLAVVGGGPAG--LAVAQVSEAGLSVCSIDNPXLIWPNNGYGVWVDFEFAMDLDDCLDA 142

Y 46 TWS-----FHEEDLTNQHEWIAPLVWHVHPDYQVFPQRRHVNNGYCVTSRHFAGILR 101

b 243 TWSGAAYIDDNTAKDLR-----PYGRVNRKQLKSNQWQ--KCIWN-----182

Y 102 QDFGQHLWLHTAVSAVHBSVQLA---DGRITIASTVIDGRGYTP-----DSALRVGFOA 153

b 183 ---GVKHFQAKVIVHEBSKMLICNDGITIQATVLDATGFSRSLVQYDKPYNPGVQV 239

Y 154 FICQEWOLSA-PHGLSSPIIMD-----ATVQQNGY--RFVYTLPLSATALLIEDT 201

b 240 AYGLIAEVEEHPFDVNMKVMFMDWRDLSHKNNTDLKERNRIPTFLYAMPFSSNRIFLEET 299

Y 202 HYIDKANIQAERARQIRDYAARQWPLQTLIRERQ-----GALPITLTDGNDQFQWQQ 255

b 300 SILVAPGLRIDDIQERWVARNLHGLIKVKSIEEDHCILPMGGPLPV-----347

Y 256 PQACSGL--RAGLFHTPTGYSLPLAVALADRL--SALDVFTSSVHTQITIAHFAQQRW-----308

b 348 PQRVVGIGGTAGVHPSTGYVWARTLAAAPVAVANAIQYLGSRSHSGNELSTAVWKDLW 407

Y 309 -----QQQFFERM-LNRMLFLAGAESRWVQRYGYPEDLIARFYAGKIT 354

b 408 PIERRQREFFCGMDILKDLPA-----TRFFDAFFOLEPRYMHGFLS 453

RESULT 11

robable glf protein - Mycobacterium tuberculosis (strain H37RV)

;Species: Mycobacterium tuberculosis

;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

;Accession: E70888

;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.

;Aandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

;Reference number: A70500; MUID:98295987; PMID:9634230

;Accession: E70888

;Status: preliminary; nucleic acid sequence not shown; translation not shown

;Molecule type: DNA

;Residues: 1-399 <COL>

;Cross-references: GB:AL022076; GB:AL123456; NID:g3256026; PIDN:CAAL7873.1; PID:g295042

;Experimental source: strain H37RV

;Genetics:

;Gene: glf

;Superfamily: Methanobacterium thermoautotrophicum UDP-galactopyranose mutase

Query Match 6.3%; Score 127.5; DB 1; Length 399;

Best Local Similarity 27.6%; Pred. No. 0.0068;

Matches 83; Conservative 22; Mismatches 91; Indels 105; Gaps 18;

Db 47 -----AYSEABPO-----TGIEVHKYGAHLF-HTSKRV 74

QY 118 HABSVLQADGRIIHASTVIDGRGYTPDSALRVGFOAQFIQEQWQLSAPHGLSPIMDATV 177

Db 75 WDYVRQFTD-----FT-DYRHRV-FAMENQAYQF--PMGL-----GLV 109

QY 178 DQONGYRFVYTLPLSATALLIEDTHVIDKANLQ--AERA-----QNIRDYAAARQ 226

Db 110 SQPFG---KYFTPEQARQIARQAABIDTADQNLEKALISLIGRPLRYFAFKVGYTAKO- 165

QY 227 WPLQTLREEQGA---LPITLTGDNQFQWQQPQACSGLRAGLFHFTTGYSLPLAVALA 282

Db 166 W--QTDPKELPAANITRLPVRVTFDNRYF-----SDTYEGL--PTDGYTAMLNMAA 213

QY 283 D 283

Db 214 D 214

RESULT 12

AD0584

Probable monooxygenase yleB [imported] - Salmonella enterica subsp. enterica serovar Typh

C;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C;Accession: AD0584

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moutie, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AD0584

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-391 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD05142.1; PID:G16501915; GSPDB:GN00176

C;Genetics:

A;Gene: yleB

C;Superfamily: ubiH protein

Query Match 5.8%; Score 117.5; DB 2; Length 391;

Best Local Similarity 23.4%; Pred. No. 0.045;

Matches 97; Conservative 42; Mismatches 163; Indels 113; Gaps 20;

QY 6 DLIVGAGLANGLIARLQQQ-----HPDMRILLIEAGPEA--CGN 44

Db 7 ETAVGGGVMGAGALAGLAQQGFTVWVIEHAPAPFVADSQPDVRIASASVALLKGL 66

QY 45 HTWSFHEEDLTINQHRWTAPLVVHHWPDYQVRFPPORRHVNSGYCVTSRHFAGILRQOF 104

Db 67 GYW---EAVQGNRSHPY--RRLTWEWENAHVFDAAELKPLLGVMVEN---NVLQAL 118

QY 105 GQHLMLHTAV-----SAVHAE---SVQLADGRIIHASTVT--DGRGYTPDSALRVG 150

Db 119 WQALAHAPGVTLRVFASLAALHRRHDGVALEADGEWTPKLVIGADGANSQVQMAGIG 178

QY 151 FOAFITQEWQLSAPHGLSSPIIMDATVDQONG-----YRFVYTLPLSATALLIEDTHYI 204

Db 179 IHA-----WQYAQ-----CMLITVKENAPGDSWQQTPTGPRAPFLP-----FD 220

QY 205 DKANLOAERARQNTIDYARQWNPLOTLLREE-----QGALPITLTGDNQRFW 252

Db 221 DWASLVWYDAPARIQOLQSLSWTQIQVEINQHFHARGVAVMVAAGAPFLT-----RRHA 275

QY 253 QQQPOQ---ACSELRAGLFHTTGYSLPLAVALADRLSALDVFTSSVH--QTIAHFAQOR 307

Db 276 LQYAPGLAVGDAHTHTPLAGQGVNIGYRDVAL--IDVLASARSVGESWASHSVLKR 333

QY 308 WQQQGFMRMLNMLEFLAGPESRWVRFGLPGLDILARFYAGKITVTDRLRL 362

Db 334 YQTR---SMADNFMWQSG-----MDL---FYAGFSLNPLPRLIL 366

Y 47 WGFHEEDLTINQHRWI-----APLVV-----HWP----- 71
b 63 LGLDENDFLAATHGVIKLGIDFVDMHTVQOSVLIHAFGPICRPLGMAPPHYWLRRQAGH 122
Y 72 -----DYQV-----RPPQRRHVNNGYY--CVTSRHFAGILRQOF----- 104
D 123 DESLWDYSINARAAGKAGRFDRIPO---VGGYLEGLAYAFHFDAAALYAQVLRYSQAQG 178
Y 105 -----GOHLWLHTAVSAVHABSVQLADGRIIHASTVID--GRGYTPDSALRVGFQ--A 153
D 179 VTRVEGRIVGVDDAEGSFVSLTLQDGKKGCGEFPIDCSGFRGLLIEEALKSGYEDWRA 238
Y 154 FIGQEWQLSAPHGLSSPII-MDATVDQMGYRFVYVTLPLSATALLIEDTHYIDKANLQAE 212
D 239 MLPMDSAIAVPCASVTPLTPYTRATAREAGWQ--WRIPLQHRT--GNGHVECSDIYIDAA 293
Y 213 RARQNIIRDYARQGWPLQTLLEEQGALPITLTGDNQFQWQOQFQACSGLRAGLFHPTTG 272
D 294 RATDVL--MANLDGAPL-----AEPQLRFT--TGRKQFWRNRVVA-MGLASGFMEPLES 344
Y 273 YSLPLAVALADRLSALDVFTSSVHQ-TIAHFAQQRWQOQGFEMLRMLFLAGPAESR- 330
D 345 TSIHLVQSALNELLAL--FPDRGFNPICIAEYNRQTALLEYEYIRDELVLHYKATTREDTP 402
Y 331 WRVMQRFYGLPEDLIARF-----YAGKLTVTDR-----LRILSGKPPVP-VFAALQA 376
D 403 FWRVAVQW-DMPDSLKARIALFSQSGRLFNQDDDLFKEASWIOVLIGGVIPQAFHPLTG 461
Y 377 IMT 379
D 462 VVT 464

arch completed: February 29, 2004, 14:52:50
b time : 15.6328 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:51:24 ; Search time 25.3626 Seconds
(without alignments)
3130.293 Million cell updates/sec

itle: US-09-941-947A-30

fect score: 2021

quence: 1 MOPYDILVAGLANGLIALRQOQHPDNRILLIRAGPAGGNTWSFHEEDLTNQR 382

oring table: BLOSUM62

atched: 809742 segs, 211153259 residues

atched: Gapop 10.0 , Gapext 0.5

al number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	2021	100.0	382	10	US-09-941-947A-30
2	2021	100.0	382	14	US-10-218-118-6
3	783.5	38.8	382	9	US-09-547-267-7
4	783.5	38.8	382	10	US-09-920-923-5
5	154.5	7.6	376	14	US-10-128-713A-20
6	152	7.5	526	9	US-09-323-998D-53
7	148.5	7.3	501	9	US-09-323-998D-55
8	142.5	7.1	502	9	US-09-323-998D-56
9	141.5	7.0	503	9	US-09-323-998D-61
10	140	6.9	500	9	US-09-323-998D-59
11	139.5	6.9	511	9	US-09-323-998D-60
12	139	6.9	498	9	US-09-323-998D-57
13	139	6.9	530	15	US-10-335-846-6
14	134.5	6.7	516	9	US-09-323-998D-54
15	133	6.6	500	9	US-09-323-998D-58

Sequence 20, Appl
Sequence 51, Appl
Sequence 5, Appl
Sequence 23, Appl
Sequence 50, Appl
Sequence 47, Appl
Sequence 26, Appl
Sequence 3, Appl
Sequence 1232, A
Sequence 2, Appl
Sequence 21, Appl
Sequence 49, Appl
Sequence 8560, Ap
Sequence 12003, A
Sequence 52, Appl
Sequence 2, Appl
Sequence 25, Appl
Sequence 48, Appl
Sequence 13043, A
Sequence 8840, Ap
Sequence 27, Appl
Sequence 4489, Ap
Sequence 14057, A
Sequence 8419, Ap
Sequence 7248, Ap
Sequence 4784, Ap
Sequence 7544, Ap
Sequence 60, Appl
Sequence 12364, A
Sequence 21815, A

16 132.5 6.6 456 9 US-09-323-998D-20
17 130.5 6.5 529 9 US-09-323-998D-51
18 130.5 6.5 529 15 US-10-335-846-5
19 129.5 6.4 529 9 US-09-323-998D-23
20 129.5 6.4 529 9 US-09-323-998D-50
21 127.5 6.3 529 9 US-09-323-998D-47
22 127 6.3 533 9 US-09-323-998D-26
23 125.5 6.2 534 15 US-10-335-846-3
24 122.5 6.1 520 15 US-10-369-493-12232
25 122.5 6.1 524 9 US-09-323-998D-2
26 122.5 6.1 524 9 US-09-323-998D-21
27 122.5 6.1 524 9 US-09-323-998D-49
28 119.5 5.9 397 14 US-10-156-761-8560
29 115 5.7 436 15 US-10-369-493-12003
30 115 5.7 533 9 US-09-323-998D-52
31 115 5.7 533 15 US-10-335-846-2
32 113.5 5.6 378 9 US-09-323-998D-25
33 113.5 5.6 378 9 US-09-323-998D-48
34 113 5.6 536 15 US-10-369-493-13043
35 112 5.5 529 15 US-10-369-493-8840
36 109.5 5.4 374 9 US-09-323-998D-27
37 108.5 5.4 537 15 US-10-369-493-4489
38 107 5.3 538 15 US-10-369-493-14057
39 106.5 5.3 374 14 US-10-156-761-8419
40 106.5 5.3 529 15 US-10-369-493-7248
41 103.5 5.1 524 15 US-10-369-493-4784
42 103.5 5.1 524 15 US-10-369-493-7544
43 102.5 5.1 933 13 US-10-001-189-60
44 101 5.0 426 15 US-10-369-493-12364
45 100 4.9 528 15 US-10-369-493-21815

ALIGNMENTS

RESULT 1

US-09-941-947A-30
; Sequence 30, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odum, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 30
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-30

Query Match 100.0%; Score 2021; DB 10; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.9e+199;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPYDILVAGLANGLIALRQOQHPDNRILLIRAGPAGGNTWSFHEEDLTNQR 60
DB 1 MOPYDILVAGLANGLIALRQOQHPDNRILLIRAGPAGGNTWSFHEEDLTNQR 60
QY 61 WIAPLVHHPDQVQFPQRRRHVNSGYCVTSRHFAGILRQOFGQHLMLHTAVSAVHAE 120

Db 61 WIAFLVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120
2y 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
Db 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
2y 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGMPLQTLRREOGAL 240
Db 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGMPLQTLRREOGAL 240
2y 241 PITLTGDNROFWOQPOACSLRAGLPHPTTGYSLPLAVLADRLSALDVFTSSSVHQT 300
Db 241 PITLTGDNROFWOQPOACSLRAGLPHPTTGYSLPLAVLADRLSALDVFTSSSVHQT 300
2y 301 AHFAQQRWQOQGFRRMLNRLFLAGPAESRWRVWQRFYGLPEDLIARFYAGKLTVTDRLR 360
Db 301 AHFAQQRWQOQGFRRMLNRLFLAGPAESRWRVWQRFYGLPEDLIARFYAGKLTVTDRLR 360
2y 361 ILSGKPPVPVFAALQAIMTTHR 382
Db 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 2

JS-10-218-118-6
; Sequence 6, Application US/10218118
; Publication No. US20030148319A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia
; APPLICANT: Rouviere, Pierre
; APPLICANT: Picataggio, Stephen
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
; FILE REFERENCE: CL1876 US NA
; CURRENT APPLICATION NUMBER: US/10/218,118
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 60/312,646
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; TYPE: PRT
; ORGANISM: Pantoea stewartii
JS-10-218-118-6

Query Match 100.0%; Score 2021; DB 14; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.9e-199;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 MOPHYDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHR 60
Db 1 MOPHYDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHR 60
2y 61 WIAFLVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120
Db 61 WIAFLVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 120
2y 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
Db 121 SVQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 180
2y 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGMPLQTLRREOGAL 240
Db 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGMPLQTLRREOGAL 240
2y 241 PITLTGDNROFWOQPOACSLRAGLPHPTTGYSLPLAVLADRLSALDVFTSSSVHQT 300
Db 241 PITLTGDNROFWOQPOACSLRAGLPHPTTGYSLPLAVLADRLSALDVFTSSSVHQT 300
2y 301 AHFAQQRWQOQGFRRMLNRLFLAGPAESRWRVWQRFYGLPEDLIARFYAGKLTVTDRLR 360

Db 301 AHFAQQRWQOQGFRRMLNRLFLAGPAESRWRVWQRFYGLPEDLIARFYAGKLTVTDRLR 360
Qy 361 ILSGKPPVPVFAALQAIMTTHR 382
Db 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 3

US-09-547-267-7
; Sequence 7, Application US/09547267
; Patent No. US20020147371A1
; GENERAL INFORMATION:
; APPLICANT: Hohmann, Hans-Peter
; APPLICANT: Passarontes, Luis
; APPLICANT: Tessier, Michel
; APPLICANT: van Loon, Adolphus
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: NJ
; COUNTRY: USA
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/547,267
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,645
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pokras, Bruce A.
; REGISTRATION NUMBER: 32,748
; REFERENCE/DOCKET NUMBER: RAN 6002/170
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-5801
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-547-267-7

Query Match 38.8%; Score 783.5; DB 9; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.6e-72;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
Qy 5 YDLILVAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTINQHWIA- 63
Db 3 HDLLIAGAGLSGALLIALAVDRPRDARIVMDARSGPSDQHTWSCHDTDLS---PSWLAR 59
Qy 64 --PLVVHWPDYQVRFPQRRHNSGYCVTSRHFAGILRQFGQHLMLHTAVSAVHAE 121
Db 60 LSPIRRGWTDQEVAFDPHSHRRLLTTGYGSTEAGALIGLLQ---GVDLRNHTVATDDTG 116
Qy 122 VQLADGRIIHASTVIDGRGVTTPDSALRVGFQAFIGQEWLSAPHLSSPINDAVDQO 181
Db 117 ATLTDGSRERACVIDARGAVETPHLTVGFKVGVETIEDAPHGVERPMDATVPMD 176
Qy 182 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIRDYAAARQGMPLQTLRREOGALP 241
Db 177 GYRFYLLPSPFRIILIEDTRYSDGSLDDGALAAQASLDYAAARGMTQGF-MRRERGILP 235

242 ITLTGDNRFQWQOQAC--SGLRAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 299
236 IALAHDAICFWEDHAQAGAVPVLGAGLPHPTTGYSLPAAQVADAIARDL--TTASARRA 294
300 IAHFAQOQWQOQGFPMNLNMLFLAGPAESRWVQRFYGLPDLIARFYAGKLTVDRL 359
295 VRGWAIDRADRDRLNMLNMLFRGCPDPRYRLQRFYLPQPLIERFYAGRLTLADRL 354
360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPILPSQAVRCL 372

RESULT 4
US-09-920-923-5
Sequence 5, Application US/09920923
Publication No. US2003002273A1
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/09/920,923
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 382
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923-5

Query Match 38.8%; Score 783.5; DB 10; Length 382;
Best Local Similarity 43.7%; Pred. No. 7.6e-72;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

2Y 5 YDLIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTLNQHRWIA- 63
DB 3 HDLLIAGAGLSGALIALAVRRDPARIVMLDARSGPSDQHTWSCHTDLS---PEWLAR 59
2Y 64 --PLAVVHPDYQVRFPPORRHVNSGYCVTSRHFAGILRQOFGQHLMLHTAVSAVHES 121
DB 60 LSPIERGENTQEVAFDHPDSRLTGTGSTEAGALIGLQ---GVDLRWNTHVATLDDTG 116
2Y 122 VOLADRIIHAHVTDGRGYPDSALRVGFOAFIQEWQLSAPHGLSSPIIMDATVDQON 181
DB 117 ATLTDGSRIEAACVIDARGAVETPHLTGVSQKFGVGEIETDAPHGVERPMIMDATVPQMD 176
2Y 182 GYRFVYTLPLSATALLIEDTHYIDKANLQARARONIRDYAARQGHPLQTLRBEQALP 241
DB 177 GYRIYLLPSPFTRILLIEDTRYSGGDLDDGALAQSLDYAARGMTGQE-MRRERGILP 235
2Y 242 ITLTGDNRFQWQOQAC--SGLRAGLPHPTTGYSLPLAVALADLSALDVTSSVHOT 299
DB 236 IALAHDAICFWEDHAQAGAVPVLGAGLPHPTTGYSLPAAQVADAIARDL--TTASARRA 294
300 IAHFAQOQWQOQGFPMNLNMLFLAGPAESRWVQRFYGLPDLIARFYAGKLTVDRL 359
295 VRGWAIDRADRDRLNMLNMLFRGCPDPRYRLQRFYLPQPLIERFYAGRLTLADRL 354
360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPILPSQAVRCL 372

RESULT 5
US-10-128-713A-20
Sequence 20, Application US/10128713A
Publication No. US20030170847A1
GENERAL INFORMATION:
APPLICANT: Bramucci, Michael G

TITLE OF INVENTION: Genes Involved in Isoprenoid Compound Production
FILE REFERENCE: CL-1788
CURRENT APPLICATION NUMBER: US/10/128,713A
CURRENT FILING DATE: 2002-04-22
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Microsoft Office 97
SEQ ID NO 20
LENGTH: 376
TYPE: PRT
ORGANISM: Rhodococcus erythropolis
US-10-128-713A-20

Query Match 7.6%; Score 154.5; DB 14; Length 376;
Best Local Similarity 23.8%; Pred. No. 4.3e-07;
Matches 75; Conservative 46; Mismatches 119; Indels 75; Gaps 15;

QY 6 DULIVGAGLANGLIARLQOQHPDMRILLIEAGPEAGGNHTWSFHEEDLTLNQHRWIAPL 65
DB 9 DVIVVGGSPAGRALATRCIAR--QLTVVVVDPHPRVWVTPTYSVWADLP---SWL--- 59
QY 66 VVHWPDYQVRFPPORRHV-NSG-----YYCVTSRHFAGILRQOFGQHLMLHTAV--- 114
DB 60 -----PDEVIASRIERPVSWTSGQKLDRIYCV-----LNTSLQSFSLSHTSIKVRG 106
QY 115 ---SAVHAESVOLADRIIHAHVTDGRGYPDSALRVGFOAFIQEWQLSAPHGLSSPI 171
DB 107 ERAQTLSTTTVVCVDSGLTGSVWVDAG--TDLAVTTAQTATG-----MIVDRALADPI 160
QY 172 I-----MDATVDOONGY-----RVVYTLPLSATALLIEDTHYIDKANLQARARQNI 218
DB 161 LGGSEAFMDWRTD--NGTSDATPSELYAVPLDDERVLLEETCLVGRPAL-----GL 211
QY 219 RDYAAARQWPLQTLRBEQALPITLTGDNRFQWQOQAC-----GLRAGLPHPTT 271
DB 212 RELETR---LRLRLHNGCEVPDDAPVERVRFAVEGRDSSPDGVLRFGGGLMHPGT 267
QY 272 GYSLPLAVALADRLS 286
DB 268 GYSVASSIAEADTWA 282

RESULT 6
US-09-323-998D-53
Sequence 53, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: SUN ZAIREN
APPLICANT: CUNNINGHAM JR., FRANCIS X.
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
CURRENT FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/624,125
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 53
LENGTH: 526
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-323-998D-53

Query Match 7.5%; Score 152; DB 9; Length 526;
Best Local Similarity 21.1%; Pred. No. 1.3e-06;
Matches 87; Conservative 66; Mismatches 167; Indels 92; Gaps 17;

QY 6 DLILVAGLANGLIALRLQOQHPDMRILLIAGPE---AGGNHTWSPHEEDLTINQHRWI 62
DB 107 DLVWICGGPAG--LALAAESAKLGNVGLV--GPDLPFTNNYGVWEDFCKDLGLQ-----157
QY 63 APLVWHVHPDYQVRPQRREHNSGVYCVTSRHP-----AGLLRQOQFQCHLWLH 111
DB 158 -ACIEHAWEDTTIYLDDBEPILIGRAYGRVSRHFLHBECLKRCVEAGV-----LYLN 208
QY 112 T-----AVSAVHAESVOLADGRII---HASTVIDGRG-----YTPDSALRVGFQAFIGOE 158
DB 209 SKVDRIEATNQSLSVECEGDVVIPCRFVTVASGAASKFLQYELGSP-RVSVQYAVGVE 267
QY 159 WQL-SAPHGLSGPIIMDATVDQONGY-----FYVTILPLSATALLIEDT 201
DB 268 VVDNPNFPDPSLMVMD-----YRDLRHAQSLEAKPTFLYAMPSPTRVFEEET 319
QY 202 HYIDKANLOAERARQIRDYAARQGWPLQTLRLREEQ-----GALPITLTGDNRFQWQOQOACSG 261
DB 320 CLASKDAMPFDLLKKXLMRLNTLGVRIKEIYEEMSVIPVGGSLPN-----TEQKTLAFG 375
QY 262 LRAGLFHPTTGYSLPLAVALADRLS-----DVTSSSVHQTIAH-----PAQ 305
DB 376 AAASVHPATGYSVRSLSSEAPKCSVLNLRQHYXNMULTSSSIPSISTQAWNTLMPQ 435
QY 306 ORWQOQGFPEMLNMLFLAGPAESRWVQORFYGLPEGLIARFYAGKLTVD 357
DB 436 ERRKRSFP-LFGLALILQLDIEGIRSPFRAPFRVYKMKWQOFLGSSLSAD 486

RESULT 7

US-09-323-998D-55
; Sequence 55, Application US/091233998D
; Patent No. US20020102631A1
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998D
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 08/937,155
; PRIOR FILING DATE: 1997-09-25
; PRIOR APPLICATION NUMBER: 08/624,125
; PRIOR FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 55
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-323-998D-55

Query Match 7.3%; Score 148.5; DB 9; Length 501;
Best Local Similarity 21.0%; Pred. No. 2.7e-06;
Matches 92; Conservative 56; Mismatches 175; Indels 105; Gaps 20;

QY 6 DLILVAGLANGLIALRLQOQHPDMRILLIAGPE---AGGNHTWSPHEEDLTINQHRWI 59
DB 84 DLVWICGGPAG--LALAAESAKLGNVGLV--GPDLPFTNNYGVWEDFCKDLGLQ-----157
QY 63 APLVWHVHPDYQVRPQRREHNSGVYCVTSRHP-----AGLLRQOQFQCHLWLH 111
DB 158 -ACIEHAWEDTTIYLDDBEPILIGRAYGRVSRHFLHBECLKRCVEAGV-----LYLN 208
QY 112 T-----AVSAVHAESVOLADGRII---HASTVIDGRG-----YTPDSALRVGFQAFIGOE 158
DB 209 SKVDRIEATNQSLSVECEGDVVIPCRFVTVASGAASKFLQYELGSP-RVSVQYAVGVE 267
QY 159 WQL-SAPHGLSGPIIMDATVDQONGY-----FYVTILPLSATALLIEDT 201
DB 268 VVDNPNFPDPSLMVMD-----YRDLRHAQSLEAKPTFLYAMPSPTRVFEEET 319
QY 202 HYIDKANLOAERARQIRDYAARQGWPLQTLRLREEQ-----GALPITLTGDNRFQWQOQOACSG 261
DB 320 CLASKDAMPFDLLKKXLMRLNTLGVRIKEIYEEMSVIPVGGSLPN-----TEQKTLAFG 375
QY 262 LRAGLFHPTTGYSLPLAVALADRLS-----DVTSSSVHQTIAH-----PAQ 305
DB 376 AAASVHPATGYSVRSLSSEAPKCSVLNLRQHYXNMULTSSSIPSISTQAWNTLMPQ 435
QY 306 ORWQOQGFPEMLNMLFLAGPAESRWVQORFYGLPEGLIARFYAGKLTVD 357
DB 436 ERRKRSFP-LFGLALILQLDIEGIRSPFRAPFRVYKMKWQOFLGSSLSAD 486

QY 161 LSA-PHGLSSPIIMDATVDQONGY-----RFVYTLPLSATALLIEDTHYIDKAN 208
DB 246 VDHHPDVPDKVFMWRDKHLDSYPCLKERNSKIPTFLYAMPFSSNRIFLSETSIVARPG 305
QY 209 LQAEARQIRDYAARQGWPLQTLRLREEQ-----GALPITLTGDNRFQWQOQOACSG 262
DB 306 LRWEDIQERMAARLKHGLNVRKIEEDERCVPIMGGPLVL-----FORVVG 353
QY 263 --RAGLFHPTTGYSLPLAVALADRLS-----ALDVFTSSSVH--QTIHFAQORW 308
DB 354 GGTAGMWHFSTGYMVARTLAAAPIVANAI VRYLSPSSNSLRGDLSAEVWRDLWPIERR 413
QY 309 QOQOQGFPEMLNMLFLAGPAESRWVQORFYGLPE-----LIAREYACKLT 355
DB 414 ROREFF-CEGMDILKLDLDA TRFDDAFDLPQHYWGHGLSSRLFLPELLVFGLSLFSH 472
QY 356 ---TDRLRLILSGKPPVPV 370
DB 473 ASNTSRLEIMT-KGTVPL 489

RESULT 8

US-09-323-998D-56
; Sequence 56, Application US/09323998D
; Patent No. US20020102631A1
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
; FILE REFERENCE: 108172-09019
; CURRENT APPLICATION NUMBER: US/09/323,998D
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 09/088,724
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 09/088,725
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 08/937,155
; PRIOR FILING DATE: 1997-09-25
; PRIOR APPLICATION NUMBER: 08/624,125
; PRIOR FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 56
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Adonis palaeestina
US-09-323-998D-56

Query Match 7.1%; Score 142.5; DB 9; Length 502;
Best Local Similarity 21.9%; Pred. No. 1.1e-05;
Matches 70; Conservative 49; Mismatches 131; Indels 69; Gaps 15;

QY 6 DLILVAGLANGLIALRLQOQHPDMRILLIAGPE---AGGNHTWSPHEEDLTINQHRWI 59
DB 87 DLVWICGGPAG--LALAAESAKLGNVGLV--GPDLPFTNNYGVWEDFCKDLGLQ-----157
QY 63 APLVWHVHPDYQVRPQR-----RRHNSGVY---CVTSRHPAGILRQOQFQCHLWLH 111
DB 145 TWSGAVV--YTDNKKYLDLPYGRVNRKQLSKMLQKCVTN-----GVKFEQA 191
QY 112 TAVSAVHAESVOLADGRII---HASTVIDGRGYP-----DSALRVGFQAFIGOE 163
DB 192 KVKIVHEESKSLICNDGITINATVLDATGFSRCLVQYDKPNPGVQVAYGMAEVEE 251
QY 164 -PHGLSSPIIMDATVDQONGY-----RFVYTLPLSATALLIEDTHYIDKAN 211
DB 252 HPFDLDKMLFMDWRSHLNEKLEKDKNRKIPTFLYAMPFSSTKIFLETSIVARPG 311
QY 212 ERARQIRDYAARQGWPLQTLRLREEQ-----GALPITLTGDNRFQWQOQOACSG 263
DB 312 EDIQERMAARLKHGLIKVKSIEEDERCVPIMGGPLVL-----FORVVG 359

Y 264 AGLEHPTTCYSLPLAVALA 282
||:||||:|
b 360 AGWHPSTGYWVARTLAAA 378
||:||||:|

RESULT 9

S-09-323-998D-61
Sequence 61, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998D

PRIOR FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1997-09-25

PRIOR APPLICATION NUMBER: 08/624,125

PRIOR FILING DATE: 1996-03-29

NUMBER OF SEQ ID NOS: 61

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 61

LENGTH: 503

TYPE: PRT

ORGANISM: Narcissus sp.

S-09-323-998D-61

Query Match

Best Local Similarity 7.0%; Score 141.5; DB 9; Length 503;

Matches 85; Conservative 66; Mismatches 150; Indels 115; Gaps 19;

6 DLILVAG-----LANGHIALRLOQHPDMRILLIEAGPEAGGHNHTWSFHEEDLTL 56

89 DLAVVGGGLARSCTSLGGGLSVSVID---PNPKLI----WPNYNG--VWVDFEDMDL 139

57 ---NOHRWIAPIVHHWPDYQVRFPPQRHRHNSGYCVTSRHFAGILRQPF----- 104

140 LCLDWTASGAIV-----YVDRSTKSLSPYARVNRKLSKMMKKCV 183

105 ---GQHLWHTAVSAVHARS---VQLADGRITHASTVIDRGYTP-----DSALRVGQAF 154

184 SNGVRFHQATVYKAKHBEKSYLICSDGVTIDARVYLDATGFSRLQYDKYPNGYQVA 243

155 IQQEWQLSA-PHGLSSPIIMD-----ATVDOONGY--RFVYTLPLGATALLIEDTH 202

244 YGILAEVEHPFDVDMKVPMDWRDHLNGKAEILNERNAKIPTFLYAMPFSSNRIPLETS 303

203 YTDKANLAQARQNRIDYAAARQGWPLQTLRREQ-----GALPITLTGDNRFQWQQP 256

304 LVARPLKMDIQERWVARLNHLGIRIKSIEEDERCVPIMGGLPVI-----P 351

257 QACSGI--RAGLFHPTTCYSPLAVALADRLSALDVTSSVHTIAH----- 302

352 QRVGIGGTAGVHPSTGY-----MVARTLAAPIVANSIVQVLYSDSLGSLGNDLSADV 405

303 ----FAQORWQQGFRRMLNRLFLAGPAERWRVMQRFYGLPDLIARFYAGKLT 354

406 WKDLWPIERRQRREFF-CFGMDILLKLDLGTTRFFDAFF----DLSPRYWHGFLS 456

RESULT 10

S-09-323-998D-59

Sequence 59, Application US/09323998D

Patent No. US20020102631A1

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998D

PRIOR FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 08/937,155

PRIOR FILING DATE: 1997-09-25

PRIOR APPLICATION NUMBER: 08/624,125

PRIOR FILING DATE: 1996-03-29

NUMBER OF SEQ ID NOS: 61

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 59

LENGTH: 500

TYPE: PRT

ORGANISM: Nicotiana sp.

US-09-323-998D-59

Query Match

Best Local Similarity 6.9%; Score 140; DB 9; Length 500;

Matches 83; Conservative 72; Mismatches 152; Indels 104; Gaps 20;

6 DLILVAGLANGHIALRLOQHPDMRILLIEAGPE---AGGNHTW--SFHEED-LTLNQH 59

85 DLAVVGGGPAG--LAVAQGVSEAGLSVVSIDPSFKLIWPNNYGVWVDFEAMDLDDCLDA 142

60 RWIAPLVHHWPDYQVRFPPQRHRHNSGYCVTSRHFAGILRQPFQOHLW-----L 110

143 TWGTVV-----YDDNTTKDLDPYGVNRKLSKMMQKCLNGVKF 186

111 HTA--VSVAHSAESVOLA---DCRIIHAFTVIDRGYTP-----DSALRVGQAFQGBWQ 160

187 HAAKVIKVIHESAKMLICNDGVTIQTWLDATGFSRLQYDKYPKPGYQVAYGILAE 246

161 LSA-PHGLSSPIIMD-----ATVDOONGY--RFVYTLPLSNTALLIEDTHYIDKAN 208

247 VEEHPFDTSKVLMWDRDHLNGKAEILNERNAKIPTFLYAMPFSSNKIFLETSILVAPG 306

209 LQAEARQNRIDYAAARQGWPLQTLRREQ-----GALPITLTGDNRFQWQQPQAC--S 260

307 LRMDDIQERWVARLNHLGIRIKSIEEDERCVPIMGGLPVI-----PQRVVG 354

261 GLRAGLFHPTTCYSPLAVALADRLSALDVTSSVH-----QTAHFAQQRW- 308

355 GGTAGLVHPSTGY-----MVARTLAAAPVAVANAIHYLGSEKDLGLNLSAAVWKLWLP 408

309 ----QOQGFPRM-LNRLFLAGPAERWRVMQRFYGLPDLIARFYAGKLT 354

409 IERRQRREFFCFGMDILLKLDLPA-----TERFFDAFFDLSPRYWHGFLS 453

RESULT 11

US-09-323-998D-60

Sequence 60, Application US/09323998D

Patent No. US20020102631A1

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND

METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998D

PRIOR FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 08/937,155

PRIOR FILING DATE: 1997-09-25

PRIOR APPLICATION NUMBER: 08/624,125


```

335 KRKLMSRLKTIQVTKYBESWYIPVGSPLN-----TEQKYLAAAGASWVHATGVS 390
275 LPLAVALADRL-----SALDVFTSSSVHQTIAHFAQORW-----QQQGF 313
391 VYRSLEAPKYASVIKILQDNAYVVSQSSA-----VNISMCAWSSLWPXKRRQRAF 446
314 FEMLRMLFLAGPAESRWVRVQFVGLPEDLIARFYAGKLTVD 357
447 PTEFGLLEIVOLDIATRTFTFRFLPTWMMWGLGSSLSFPD 490

SULT 14
-09-323-998D-54
Sequence 54, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1996-03-29
PRIOR APPLICATION NUMBER: 08/624,125
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 54
LENGTH: 516
TYPE: PRI
ORGANISM: Tagetes erecta
S-09-323-998D-54

Query Match 6.7%; Score 134.5; DB 9; Length 516;
Best Local Similarity 19.9%; Pred. No. 7.9e-05;
Matches 86; Conservative 65; Mismatches 188; Indels 93; Gaps 17;

Y 6 DLILVAGLGLALRLQOQHDPDRILLTEAGPEAGGKHVTSFEEDLTINQHWIAPL 65
b 96 DLVVGCGPAG--LALAGESAKGLINALL--GFDLPFTNNYGVWEDEFI---GLGLEGC 148
Y 66 VHWHPDYQVRRPQRRHNSGYCVTSRHFAGILRQOFGQHLMLHTAVSAVHAESVOL- 124
b 149 IEHVNRDTVV-----YLDNDPILIGRAYGRVSRDLLHEELLTRCMESGVSYLSKVE 201
Y 125 -----ADGRI---IHASTVIDG--RGYTPDSAL---RVGFQARTIGQEWOL-S 162
b 202 RITEAPNGLSLICEGNTIPCRLATVAGSAAGSKLLQYELGGPRVCVQTAYGIEVEVES 261
Y 163 APHGLSSPIINDATVDOONGYR-----FVYTLPLSATALLIEDTHYIDK 206
b 262 IEPDPSLVFMD-----YRDYTKHSQSLEAQVPTFLYVWPMSPTKVFFETCLASK 313
Y 207 ANLQAEARQWIRDYAARQGWPLQTLRLREQGALPITLTGDNRQFWQOQPCAGSLRAGL 266
b 314 EAMPFELLKTKLMSRLKTIQVTKYBESWYIPVGSPLN-----TEQKYLAAAGASW 369
Y 267 FHPTTGYSLPLAVALADRLSAL--DVFTSSSVHQTIAHFAQORW----- 309
b 370 VHPATGYSVRSVRSSEAPNYAAVIAKILGKNSQKMLDHQRYTTNISKQAWETLPLERKR 429
Y 310 QQGFPRMLNMLFLAGPAESRWVRVQFVGLPEDLIARFYAGKLTVTORLILSGXDPVP 369
b 430 QRAFF-LFGLALIVQMDIEGTRTFRTFFELPTWMMWGLGSSLSSTOLI----- 478
Y 370 VPAALQAIMTTH 381

```

```

Db 479 IFATWFIAPH 490

RESULT 15
US-09-323-998D-58
Sequence 58, Application US/09323998D
Patent No. US20020102631A1
GENERAL INFORMATION:
APPLICANT: CUNNINGHAM JR., FRANCIS X.
APPLICANT: SUN, ZAIREN
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND
METHODS OF USE THEREOF
FILE REFERENCE: 108172-09019
CURRENT APPLICATION NUMBER: US/09/323,998D
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: 09/088,724
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 09/088,725
PRIOR FILING DATE: 1998-06-02
PRIOR FILING DATE: 1997-09-25
PRIOR APPLICATION NUMBER: 08/937,155
PRIOR FILING DATE: 1996-03-29
PRIOR APPLICATION NUMBER: 08/624,125
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 58
LENGTH: 500
TYPE: PRI
ORGANISM: Lycopersicon esculentum
US-09-323-998D-58

Query Match 6.6%; Score 133; DB 9; Length 500;
Best Local Similarity 19.4%; Pred. No. 0.00011;
Matches 80; Conservative 74; Mismatches 152; Indels 106; Gaps 19;

QY 6 DLILVAGLGLALRLQOQHDPDRILLTEAGPEA--CGNH----- 45
Db 85 DLAVVGGPAG--LAVACQVSEAGLSVCSIDPDKLIWNNYGVWVDFEAMDLLDCLDA 142
QY 46 TMS-----FHEEDLTINQHWIAPLAVHWPDYQVRRPQRRHNSGYCVTSRHFAGILR 101
Db 143 TMSGAAVTIDNTAKDLHR-----PYGRVNRKQLKSKMWQ--KCMN----- 182
QY 102 QQFGQHLMLHTAVSAVHAESVOLA---DGRTHASTVIDGRGVTP-----DSALRVGFOA 153
Db 183 ---GVKFRQAKVIKVIHEESKSMICNDGITIQATVLDATGFSRSLVQYDXPNFYQV 239
QY 154 FIGQEWOLSA--PHGLSSPIIND-----ATVDOONGY--RFVYTLPLSATALLIEDT 201
Db 240 AVGILARVEEHPFDVNRVNFMDWRDLSHKNTDLKXNSRIPTFLYAMPFSSNRIPLBET 299
QY 202 HYIDKANQAERARQWIRDYAARQGWPLQTLRLREQ-----GALPITLTGDNRQFWOQ 255
Db 300 SILVAPGLRIDDIQERVARLNHLGIKVSIEDEHCLIPMGSPPLVL----- 347
QY 256 PQACSGL--RAGLFHPTTGYSLPLAVALADRL--SALDVFTSSSVHQTIAHFAQORW----- 308
Db 348 FORVVGIGGTAGWHPSTGYVARTLAAAEVANAIIQYLGSRSHSGNELSTAVWKOLW 407
QY 309 -----QQQGFPRM--LNRMLFLAGPAESRWVRVQFVGLPEDLIARFYAGKLT 354
Db 408 PIERRRQREFFCFQMGDILLKDLPA-----TRRFTDAFFDLEPRVWHGFLS 453

Search completed: February 29, 2004, 15:28:00
Job time : 27.3626 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:35:44 ; Search time 11.6083 Seconds
(without alignments)
1698.885 Million cell updates/sec

itle: US-09-941-947A-30

erfect score: 2021
equence: 1 MQPHYDLIIIVAGLANGLIA.....SGKPPVVFALQAIMTHR 382

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

ral number of hits satisfying chosen parameters: 389414

imum DB seq length: 0
iximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Issued Patents AA:*
1: /cgn2_6/ptodata/2/aa/5A.COMB.pdp:*
2: /cgn2_6/ptodata/2/aa/5B.COMB.pdp:*
3: /cgn2_6/ptodata/2/aa/6A.COMB.pdp:*
4: /cgn2_6/ptodata/2/aa/6B.COMB.pdp:*
5: /cgn2_6/ptodata/2/aa/ECTUS.COMB.pdp:*
6: /cgn2_6/ptodata/2/aa/backfiles.pdp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1678	83.0	382	1 US-07-783-705A-3	Sequence 3, Appli
2	1127	55.8	374	1 US-08-095-726-14	Sequence 14, Appl
3	1127	55.8	374	1 US-08-096-623A-14	Sequence 14, Appl
4	783.5	38.8	382	3 US-08-660-645A-7	Sequence 7, Appli
5	783.5	38.8	382	3 US-09-298-718-7	Sequence 7, Appli
6	783.5	38.8	382	3 US-09-546-969-7	Sequence 7, Appli
7	783.5	38.8	382	3 US-08-980-832-5	Sequence 5, Appli
8	783.5	38.8	382	4 US-09-547-267-7	Sequence 7, Appli
9	783.5	38.8	382	4 US-09-920-923B-5	Sequence 5, Appli
10	759.5	37.6	386	2 US-08-663-310-6	Sequence 6, Appli
11	759.5	37.6	386	2 US-09-006-491-6	Sequence 6, Appli
12	759.5	37.6	386	3 US-09-335-919-6	Sequence 6, Appli
13	154.5	7.6	498	3 US-09-134-607A-19	Sequence 19, Appl
14	152.5	7.5	498	3 US-09-134-607A-17	Sequence 17, Appl
15	152.5	7.5	498	3 US-09-134-607A-18	Sequence 18, Appl
16	152	7.5	526	4 US-09-323-998E-53	Sequence 53, Appl
17	148.5	7.3	411	1 US-08-393-561-2	Sequence 2, Appli
18	148.5	7.3	498	2 US-08-702-598-2	Sequence 2, Appli
19	148.5	7.3	501	4 US-09-323-998E-55	Sequence 55, Appl
20	142.5	7.1	502	4 US-09-323-998E-56	Sequence 56, Appl
21	141.5	7.0	503	4 US-09-323-998E-61	Sequence 61, Appl
22	140	6.9	500	4 US-09-323-998E-59	Sequence 59, Appl
23	139.5	6.9	511	3 US-09-201-641-2	Sequence 2, Appli
24	139.5	6.9	511	4 US-09-323-998E-60	Sequence 60, Appl
25	139	6.9	498	4 US-09-323-998E-57	Sequence 57, Appl
26	134.5	6.7	516	3 US-09-201-641-6	Sequence 6, Appli
27	134.5	6.7	516	4 US-09-323-998E-54	Sequence 54, Appl

Query Match 83.0%; Score 1678; DB 1; Length 382;

Sequence 58, Appl
Sequence 20, Appl
Sequence 20, Appl
Sequence 51, Appl
Sequence 23, Appl
Sequence 47, Appl
Sequence 50, Appl
Sequence 26, Appl
Sequence 2, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 21, Appl
Sequence 2, Appl
Sequence 49, Appl
Sequence 21, Appl
Sequence 6479, Ap
Sequence 52, Appl
Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-07-783-705A-3
; Sequence 3, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, Shigeyuki
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; NUMBER OF INVENTION: SYNTHESIS OF CAROTENOIDS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ladass & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/783,705A
; FILING DATE: 19911023
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-103078
; FILING DATE: 21-APR-1989
; APPLICATION NUMBER: JP 2-53225
; FILING DATE: 05-MAR-1990
; APPLICATION NUMBER: US 07/519,011
; FILING DATE: 19-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwadron, Janet I.
; REGISTRATION NUMBER: 33,778
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-708-1935
; TELEFAX: 212-246-5959
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-705A-3

Best Local Similarity 82.2%; Pred. No. 1.2e-159;
Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

QY 1 MQPHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHR 60
DB 1 MQPHYDLILVAGLANGLIALRLQOQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHR 60
QY 61 WTAPLVVHHWPDYQVRFQRRHVNNGYCVTSRHFAGILRQOQHLMHTAVSAVHA 120
DB 61 WTAPLVVHHWPDYQVRFQRRHVNNGYCVTSRHFAGILRQOQHLMHTAVSAVHA 120
QY 121 SVQLADGRIIHAFTVIGRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDDQ 180
DB 121 SVQLADGRIIHAFTVIGRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDDQ 180
QY 181 NGYRFVYTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGMPLQTLRREBQAL 240
DB 181 NGYRFVYTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGMPLQTLRREBQAL 240
QY 241 PITLGDNRQFQOQPOACSGLRAGLFHPTTGYSPLAVALADRLSALDVFTSSVHTI 300
DB 241 PITLGDNRQFQOQPOACSGLRAGLFHPTTGYSPLAVALADRLSALDVFTSSVHTI 300
QY 301 ASHPAQRWQOQGFPMNLNMLFLAGPAESRWVQRYGLPEDIARFYAGKLTVDRLR 360
DB 301 ASHPAQRWQOQGFPMNLNMLFLAGPAESRWVQRYGLPEDIARFYAGKLTVDRLR 360
QY 361 ILSGKPPVPEVFAALQAIMTTHR 382
DB 361 ILSGKPPVPEVFAALQAIMTTHR 382

RESULT 2
US-08-095-726-14
; Sequence 14, Application US/08095726
; Patent No. 5530188
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brinkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Huel-Che B.
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in
; TITLE OF INVENTION: Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept
; STREET: 200 E Randolph St
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60680-0703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,726
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/785,566
; FILING DATE: 30-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, No. 5530188val B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 3128567180
; TELEFAX: 3128564972
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-095-726-14

Query Match 55.8%; Score 1127; DB 1; Length 374;
Best Local Similarity 58.4%; Pred. No. 1.7e-104;
Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;

QY 6 DULIVGAGLANGLIALRLQOQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHRWIAPL 65
DB 3 DULIVGAGLANGLIALRLQOQHPDMRILLIEAGPEAGGNHTWSPHEEDTLNQHRWIAPL 62
QY 66 VVHWPDYQVRFQRRHVNNGYCVTSRHFAGILRQOQHLMHTAVSAVHAESVOLA 125
DB 63 VAHWPDYQVRFQRRHVNNGYCVTSRHFAGILRQOQHLMHTAVSAVHAESVOLA 122
QY 126 DGRITHASTVIDRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDDQNGYRF 185
DB 123 NGEALLAGAVIDRGYTPDSALRVGFQAFIQEWQLSAPHLSSPIIMDATVDDQNGYRF 182
QY 186 VYTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGMPLQTLRREBQALPITL 245
DB 183 VYTLPLSATALIEDTHYIDKANLQAEARONIRDYAARQGMPLQTLRREBQALPITL 242
QY 246 GDNQFQWQOQGFPMNLNMLFLAGPAESRWVQRYGLPEDIARFYAGKLTVDRLRILSGK 305
DB 243 VTSRLCGPMRREASCMRAGLFHPTTGYSPLAVALADRLSALDVFTSSVHTIHPAQ 302
QY 306 ORWQOQGFPMNLNMLFLAGPAESRWVQRYGLPEDIARFYAGKLTVDRLRILSGK 365
DB 303 RHWKQGFPMRREASCMRAGLFHPTTGYSPLAVALADRLSALDVFTSSVHTIHPAQ 362
QY 366 PVPV 370
DB 363 PVPV 367

RESULT 3
US-08-096-623A-14
; Sequence 14, Application US/08096623A
; Patent No. 5684238
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brinkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Huel-Che B.
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
; TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welsh & Katz, Ltd.
; STREET: 120 S. Riverside Plaza, 22nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,623A
; FILING DATE: 22-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,061
; FILING DATE: 09-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,921

FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA: US 07/562,674
APPLICATION NUMBER: 03-AUG-1990
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
IS-08-096-623A-14

Query Match 55.8%; Score 1127; DB 1; Length 374;
Best Local Similarity 58.4%; Pred. No. 1.7e-104;
Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;

6 DLILVGLAGLIALRLOOQHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIAPL 65
3 DLILVGLAGLIALRLOOQHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIAPL 62
66 VVHEWPDYQVRRPQRRHNSGYCVTSRHFAGILRQOQHLWHTAVSAVHAESVOLA 125
63 VAHAWPGYEQFDPRLRLRGLGYSTSRFAALHQAALGENIWNCSVSEVLPSVRLA 122
126 DGRHIIHASTVIDGRYTPDSALRVGQAFIGQEWQLSAPHGLSSPIIMDATVDQNGYRF 185
123 NGEALLAGAVIDGRYTPDSALRVGQAFIGQEWQLSAPHGLSSPIIMDATVDQNGYRF 182
186 VYTLPLSATALLIEDTHYDKANLQAEARQNIQDYAARQGNPLOTLLREOGALPITLT 245
183 VYTLPLSATALLIEDTHYDKANLQAEARQNIQDYAARQGNPLOTLLREOGALPITLT 242
246 GDNRFQWQOQOQAGLFTPTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFAQ 305
243 VTSRLCGPMERRAASGNRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHQTIAHFAQ 302
306 QRWQOQOQFEMLNMLFLAGPAESRWVMOQFYGLPDLIARFYAGKLTVDRLRLSOK 365
303 RHWROQGFPELLNMLFLAGPAESRWVMOQFYGLPDLIARFYAGKLTVDRLRLSOK 362
366 PPVPV 370
363 PPVPL 367

RESULT 4
US-08-660-645A-7
Sequence 7, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: PERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street

5 YDLILVGLAGLIALRLOOQHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIA- 63
3 HDLLIAGAGLUSGALLAVALDRRDPDARIWLDARSGDSQHTWSDTDLUS---PEWLEAR 59
64 --PLVWHPDYQVRRPQRRHNSGYCVTSRHFAGILRQOQHLWHTAVSAVHAES 121
60 LSPRRGENTDQEVAFDHSRLTTGYSGIEAGALIGLLQ---GVDLRWNTHTVATLDDTG 116
122 VOLADGRHIIHASTVIDGRYTPDSALRVGQAFIGQEWQLSAPHGLSSPIIMDATVDQON 181
117 ATLDGSRIRIACACVIDARGAVETPHLTGVGFQKGVVEIETDAPHGVERPMIMDATVPQMD 176
182 GYRFVYTLPLSATALLIEDTHYDKANLQAEARQNIQDYAARQGNPLOTLLREOGALP 241
177 GYRFYLLPSPTRILLIEDTHYDGGDLDDGALQAASLDYAARRGWTGQE-MREREGILP 235
242 ITLGTGDNRFQWQOQOQAC--SGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHQT 299
236 IALAHAIDGFWRHAQGAQVPGVGLGAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHQT 294
300 IAHEAQORWQOQOQFEMLNMLFLAGPAESRWVMOQFYGLPDLIARFYAGKLTVDRL 359
295 VGVWAIDRADRDRLNMLFLAGPAESRWVMOQFYGLPDLIARFYAGKLTVDRL 354
360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPPLSQAVRCL 372

RESULT 5
US-09-298-718-7
Sequence 7, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel

CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-660-645A-7

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

5 YDLILVGLAGLIALRLOOQHPDMRILLIEAGPEAGGNTWSPHEEDTLNCHRWIA- 63
3 HDLLIAGAGLUSGALLAVALDRRDPDARIWLDARSGDSQHTWSDTDLUS---PEWLEAR 59
64 --PLVWHPDYQVRRPQRRHNSGYCVTSRHFAGILRQOQHLWHTAVSAVHAES 121
60 LSPRRGENTDQEVAFDHSRLTTGYSGIEAGALIGLLQ---GVDLRWNTHTVATLDDTG 116
122 VOLADGRHIIHASTVIDGRYTPDSALRVGQAFIGQEWQLSAPHGLSSPIIMDATVDQON 181
117 ATLDGSRIRIACACVIDARGAVETPHLTGVGFQKGVVEIETDAPHGVERPMIMDATVPQMD 176
182 GYRFVYTLPLSATALLIEDTHYDKANLQAEARQNIQDYAARQGNPLOTLLREOGALP 241
177 GYRFYLLPSPTRILLIEDTHYDGGDLDDGALQAASLDYAARRGWTGQE-MREREGILP 235
242 ITLGTGDNRFQWQOQOQAC--SGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHQT 299
236 IALAHAIDGFWRHAQGAQVPGVGLGAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHQT 294
300 IAHEAQORWQOQOQFEMLNMLFLAGPAESRWVMOQFYGLPDLIARFYAGKLTVDRL 359
295 VGVWAIDRADRDRLNMLFLAGPAESRWVMOQFYGLPDLIARFYAGKLTVDRL 354
360 RILSGKPPVPVFAALQAI 377
355 RIVTGRPPPLSQAVRCL 372

RESULT 5
US-09-298-718-7
Sequence 7, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-298-718-7

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
QY 5 YDLILVGLAGLALRLQOQHDPMDRILLIEAGPEAGGNHTWSFHEEDLTINQHRWIA- 63
DB 3 HDLLIAGAGLSGALLAVALRRDPDARIWMLDARSQSDQHTWSCHTDLS---PEWLAR 59
QY 64 --PLVVHWPDYQVRRPQRRRHVNSGYCVTSRHFAGILRQOQGHILWLTAVSAVHAES 121
DB 60 LSPFIRRGWTDQEVAFPDHSRRLTTGYGSIIEAGALIGLLQ---GVDLRWNTHVATLDDTG 116
QY 122 VOLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHGLSSPIIMDATVDDQN 181
DB 117 ATLTDGSRIEAACVIDARGAVETPHLTGVGFQGVETETDAPHGVERPMDATVPQMD 176
QY 182 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIDYAAARQGWPLQTLLEEQGALP 241
DB 236 IALAHAICGFWRDHQAQGVAVPVGLAGLPHFVTGYSLPVAQAQVADAIAARDL-TTASARRA 294
QY 300 IAHFAQQRWQOQGFPRMLNRMFLAGPAESRWYQRYGLPEDIARFVAGKLTVDRL 359
DB 295 VRGWAIDRADRDRLRLNRMFLRGCPPDRYRLLRQFYRLPQPLIERFYAGRLTLADRL 354
QY 360 RILSGKPPVPVFAALQAI 377
DB 355 RIVTGRPPILPSQAVRCL 372

RESULT 6

US-09-546-969-7

Sequence 7, Application US/09546969
Patent No. 6207409
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/546,969
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-546-969-7

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;
QY 5 YDLILVGLAGLALRLQOQHDPMDRILLIEAGPEAGGNHTWSFHEEDLTINQHRWIA- 63
DB 3 HDLLIAGAGLSGALLAVALRRDPDARIWMLDARSQSDQHTWSCHTDLS---PEWLAR 59
QY 64 --PLVVHWPDYQVRRPQRRRHVNSGYCVTSRHFAGILRQOQGHILWLTAVSAVHAES 121
DB 60 LSPFIRRGWTDQEVAFPDHSRRLTTGYGSIIEAGALIGLLQ---GVDLRWNTHVATLDDTG 116
QY 122 VOLADGRIIHASTVIDGRGYTPDSALRVGFQAFIQEWQLSAPHGLSSPIIMDATVDDQN 181
DB 117 ATLTDGSRIEAACVIDARGAVETPHLTGVGFQGVETETDAPHGVERPMDATVPQMD 176
QY 182 GYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIDYAAARQGWPLQTLLEEQGALP 241
DB 177 GYRFYLLPFSPTRIILIEDTRYSDGDLDDGALAQAASLDYAAARGWTGQE-MRREGILP 235
QY 242 ITLTGNRQFWQOQPAQ--SGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSSVHQT 299
DB 236 IALAHAICGFWRDHQAQGVAVPVGLAGLPHFVTGYSLPVAQAQVADAIAARDL-TTASARRA 294
QY 300 IAHFAQQRWQOQGFPRMLNRMFLAGPAESRWYQRYGLPEDIARFVAGKLTVDRL 359
DB 295 VRGWAIDRADRDRLRLNRMFLRGCPPDRYRLLRQFYRLPQPLIERFYAGRLTLADRL 354
QY 360 RILSGKPPVPVFAALQAI 377

355 RIVTGRPPPLSQAVRCL 372

RESULT 7

S-08-980-832-5

Sequence 5, Application US/08980832B

Patent No. 6291204

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/08/980,832B

CURRENT FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 5

LENGTH: 382

TYPE: PRT

ORGANISM: Flavobacterium sp. R1534

S-08-980-832-5

Query Match 38.8%; Score 783.5; DB 3; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

Y 5 YDILIVGAGLIARLRCQQHPDNRILLIIEAGPEAGNHTWSPHEBDLTINQHRWIA- 63
b 3 HDLLIAGAGLSGALIALAVDRRDPDARIWMLDARSQPSDQHTWSCHDTLS----PEWLAR 59
Y 64 --PLVVHHPDYVRRPQRRRHVNSGYCVTSRHFAGILRQOFGQHLMLHTAVSAVHAES 121
b 60 LSPIRGEWTDQEVAFPDHSRRLTTGYSIEAGALIGLLQ---GVDLRNTHVATLDDTG 116
Y 122 VQADGRIIHAFTVIDGRGYTPDSALRVGFQAFIQQEWOLSPHGLSSPIIMDATVDQON 181
b 117 ATLTDGSRIEAACVIDARGAVETPHLTVGQFQVGVIEITDAPHGVVERPMIMDATVPMQMD 176
Y 182 GYEFVYTLPLSATALIETHYIDKANLQAEARQNIQRYAARQWPLQTLLEEGALP 241
b 177 GYRFFIYLLPSPRIILIEDTRYSDGGLDGGALAAQSLDYAARRGWTGQ--MRRERGILP 235
Y 242 ITLTGDNROFWOQOQOAC--SGLRAGLPHTTGYSLPLAVALADRLSLDVTSSVHQT 299
b 236 IALAHDAIGFWRDHAQGVAVPGLGAGLFPHVTGYSLPYAAQVADAIAARDL--TTASARRA 294
Y 300 IAHFAQORWQOQGFPMNMLFLAGPAESRRVWQRFYGLPEDLIARFYAGKLTVTDR 359
b 295 VRGWAIDRADRDRLFLNMLFRGCPDPRYRLQLRQFYRLPQPLIERFYAGRLTLADRL 354
Y 360 RILSGKPPVPVFAALQAI 377
b 355 RIVTGRPPPLSQAVRCL 372

RESULT 8

S-09-547-267-7

Sequence 7, Application US/09547267

Patent No. 6613543

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-547-267-7

Query Match 38.8%; Score 783.5; DB 4; Length 382;
Best Local Similarity 43.7%; Pred. No. 4.4e-70;
Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 6;

Qy 5 YDILIVGAGLIARLRCQQHPDNRILLIIEAGPEAGNHTWSPHEBDLTINQHRWIA- 63
Db 3 HDLLIAGAGLSGALIALAVDRRDPDARIWMLDARSQPSDQHTWSCHDTLS----PEWLAR 59
Qy 64 --PLVVHHPDYVRRPQRRRHVNSGYCVTSRHFAGILRQOFGQHLMLHTAVSAVHAES 121
Db 60 LSPIRGEWTDQEVAFPDHSRRLTTGYSIEAGALIGLLQ---GVDLRNTHVATLDDTG 116
Qy 122 VQADGRIIHAFTVIDGRGYTPDSALRVGFQAFIQQEWOLSPHGLSSPIIMDATVDQON 181
Db 117 ATLTDGSRIEAACVIDARGAVETPHLTVGQFQVGVIEITDAPHGVVERPMIMDATVPMQMD 176
Qy 182 GYEFVYTLPLSATALIETHYIDKANLQAEARQNIQRYAARQWPLQTLLEEGALP 241
Db 177 GYRFFIYLLPSPRIILIEDTRYSDGGLDGGALAAQSLDYAARRGWTGQ--MRRERGILP 235
Qy 242 ITLTGDNROFWOQOQOAC--SGLRAGLPHTTGYSLPLAVALADRLSLDVTSSVHQT 299
Db 236 IALAHDAIGFWRDHAQGVAVPGLGAGLFPHVTGYSLPYAAQVADAIAARDL--TTASARRA 294
Qy 300 IAHFAQORWQOQGFPMNMLFLAGPAESRRVWQRFYGLPEDLIARFYAGKLTVTDR 359
Db 295 VRGWAIDRADRDRLFLNMLFRGCPDPRYRLQLRQFYRLPQPLIERFYAGRLTLADRL 354
Qy 360 RILSGKPPVPVFAALQAI 377
Db 355 RIVTGRPPPLSQAVRCL 372

RESULT 9

US-09-920-923B-5

Sequence 5, Application US/09920923B

Patent No. 6677134

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Fermentative Carotenoid Production

FILE REFERENCE: 15464 US (C38435/125944)

CURRENT APPLICATION NUMBER: US/09/920,923B

CURRENT FILING DATE: 2001-08-02

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/006,491

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/663,310
FILING DATE: 23-SEP-1996
APPLICATION NUMBER: WO PCT/JP94/02220
FILING DATE: 26-DEC-1994

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: JP 6-235917
FILING DATE: 05-SEP-1994
PRIORITY APPLICATION DATA:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: JP 5-348737
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/109
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 386 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

S-09-006-491-6

Query Match 37.6%; Score 759.5; DB 2; Length 386;

Best Local Similarity 43.0%; Pred. No. 1.1e-67;

Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;

```
Y 5 YDLILVAGLANGLIALRLQQQHPDMRILLIE--AGPEAGNHTWSFHEEDTLNQHRWI 62
b 3 HDVLLAGAGLANGLIALALRAARPDRLVLLDHAAGPSDG--HTWSCHDPDLSPD--WL 57
Y 63 A---PLVWHWPDYQVRFPORRHVNSGYCVTSRHFAGILRQFGQHLWLHTAVSAVHA 119
b 58 ARLKPLRRANWPDQVRFPPHARLATGYSLDGALADAVVRSAGAEIRW-DSDIALLDA 116
Y 120 ESQLADGRIIHAISTVIDGRGVTTPDSALRVGFQAFIGQEWQLSAPHGLSSPTIMDATVQ 179
b 117 QGATLSCGTRIEAGVLDGRGAQPSRHLTVGFQKFGVGEIETDRPHGVPRMIMDATVQ 176
Y 180 QNGYRVFTYLPISATALLIEDTHYIDKANLQERARQNRIDYAAQGWPLQTLREBOGA 239
b 177 QDGYRIYLLPSPTRILLIEDTRYSDGDLDDALLAASHDFARQQGN-TGAEVRENGI 235
Y 240 LPITLGDNRQFQQQP--QACSGLRAGLFHPTTGYSLPLAVALADRLSALD-VFTSSV 296
b 236 LPITALHAAGAAGWADHAAGFVFGVLRAGFPHVPTGYSLPYAAQVADVAGLSGPGGTAL 295
Y 297 HQPIAHFAQRWQQQFFFMILNMLFLAGPAESRWVMOREVGILPEDLIARPYAGKLTVT 356
b 296 RGAIRDYADRAKRDREKLLNMLFRGCAPDRRTYLLQRFVPMRPHGLIBRFYAGRUSVA 355
Y 357 DRLRLSGKPPVPVFAALQAI 377
b 356 DQLRIVTGKPPPLGTAIRCL 376
```

RESULT 12

US-09-335-919-6
Sequence 6, Application US/09335919
Patent No. 6150130

GENERAL INFORMATION:

APPLICANT: MISAHA, No. 6150130hiko
APPLICANT: KONDO, Keiji
APPLICANT: KAJIMARA, Susumu

APPLICANT: YOKOYAMA, Akihiro

TITLE OF INVENTION: DNA STRANDS USEFUL FOR THE SYNTHESIS OF
XANTHOPHYLLS AND THE PROCESS FOR PRODUCING THE
XANTHOPHYLLS

TITLE OF INVENTION: XANTHOPHYLLS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/335,919

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA: US/08/663,310

FILING DATE: 23-SEP-1996

APPLICATION NUMBER: WO PCT/JP94/02220

FILING DATE: 26-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 6-235917

FILING DATE: 05-SEP-1994

PRIOR APPLICATION DATA: JP 5-348737

FILING DATE: 27-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 49441/109

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 386 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-335-919-6

Query Match 37.6%; Score 759.5; DB 3; Length 386;

Best Local Similarity 43.0%; Pred. No. 1.1e-67;

Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;

```
Qy 5 YDLILVAGLANGLIALRLQQQHPDMRILLIE--AGPEAGNHTWSFHEEDTLNQHRWI 62
Db 3 HDVLLAGAGLANGLIALALRAARPDRLVLLDHAAGPSDG--HTWSCHDPDLSPD--WL 57
Qy 63 A---PLVWHWPDYQVRFPORRHVNSGYCVTSRHFAGILRQFGQHLWLHTAVSAVHA 119
Db 58 ARLKPLRRANWPDQVRFPPHARLATGYSLDGALADAVVRSAGAEIRW-DSDIALLDA 116
Qy 120 ESQLADGRIIHAISTVIDGRGVTTPDSALRVGFQAFIGQEWQLSAPHGLSSPTIMDATVQ 179
Db 117 QGATLSCGTRIEAGVLDGRGAQPSRHLTVGFQKFGVGEIETDRPHGVPRMIMDATVQ 176
Qy 180 QNGYRVFTYLPISATALLIEDTHYIDKANLQERARQNRIDYAAQGWPLQTLREBOGA 239
```

Db 177 QDGYRFLYLLPSPTRILIEDTRYSDGGDLDDALAAASHDYARQQOW--TCAEVRREIRGI 235
QY 240 LPITUTGNORFWOQP--QACSLRAGLPHTTGYSLPLAVALADRLSALD-VFTSSSV 296
Db 236 LPIALHDAAGFADHAGPVPVGLRAGFFPVTGYSLPAAQVADVAGLSGPPGTDAL 295
QY 297 HQTIAFQACQWQOQGFPRMLNMLFLAGPAESRWVWQRFYGLPDLIARFVAGKLTUTV 356
Db 296 RGAIKDYAIDARDRFLRLNRLPRGAPDRRYTLQRFYRPHGLIERFVAGRLSVA 355
QY 357 DELRLILSGKPPVPVFAALQAI 377
Db 356 DQLRIVTGKPPPIUGTARCL 376

RESULT 13

US-09-134-607A-19
Sequence 19, Application US/09134607A

Patent No. 6252141

GENERAL INFORMATION:

APPLICANT: Joseph Hirschberg et al.

TITLE OF INVENTION: POLYNUCLEOTIDES CONTROLLING THE EXPRESSION

TITLE OF INVENTION: OF AND CODING FOR GENE B IN TOMATO AND USE

TITLE OF INVENTION: OF SAME FOR ALTERING CAROTENOID

TITLE OF INVENTION: BIOSYNTHESIS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Castorina

STREET: 20001 Jefferson Davis Highway, Suite 207

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: Twinhead, Slimnote 990TX

OPERATING SYSTEM: MS DOS version 6.2,

OPERATING SYSTEM: Windows version 3.11

SOFTWARE: Word for Windows version 2.0,

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/134.607A

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 325/12

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX:

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 498

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-134-607A-19

Query Match 7.6%; Score 154.5; DB 3; Length 498;

Best Local Similarity 22.1%; Pred. No. 6.3e-07;

Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;

QY 5 YDLILVAGLANGLIALRLOQO--HPDMRILLIAGPEA--GGNHTWSPFEEDITLN-- 57

Db 82 FDVILIGAGPA---GLRLAEQSVKYGKVCVDSPLSMFNNYGVWVDFENIGLEDC 137

QY 58 -QHRVAPLWVHWPYQVRP---PQRR-----RHVNSGYCVTSRHPAGILRQOFG 105

Db 138 LDHKV--PMTCVHINDNKTGILGRPYGRVSRKKLKLKLLNS---CVENR-----VKFYK 186
QY 106 QHLWLETAWSAVHAE---SVQLADGRIIHASTVIDGRGYTP-----DSALRVGFAQFIGQ 157
Db 187 AKVW-----KVEHEFESSIIVCDDGKIRGSLVWDASGFASDFIEYDKPRNHGYQIAHGV 241
QY 158 EQWL-SABHGSLSSPLIND-----ATVDCQNGYRFVYTLPLSATALLIEDTHYID 205
Db 242 LVEVNHDFDLDKVMVLDWRDRLSHGNPFYLVNNAKEPTFLYAMPFDRNLVFLBETSLVS 301
QY 206 KANLQAEARQIRDYAARQGMPLQTLRREEQGALPITLTGDNRFQWQOQFQACSGLRAG 265
Db 302 RVLVSMEVEXRMVRLRHILGKIVRSVIEEKCVIPM---GGPLRIPQNVMAIGG-NSG 357
QY 266 LFPHPTTGYSLPLAVALADRLS 286
Db 358 IVHPSTGYMVARSMALAEVLA 378

RESULT 14

US-09-134-607A-17

Sequence 17, Application US/09134607A

Patent No. 6252141

GENERAL INFORMATION:

APPLICANT: Joseph Hirschberg et al.

TITLE OF INVENTION: POLYNUCLEOTIDES CONTROLLING THE EXPRESSION

TITLE OF INVENTION: OF AND CODING FOR GENE B IN TOMATO AND USE

TITLE OF INVENTION: OF SAME FOR ALTERING CAROTENOID

TITLE OF INVENTION: BIOSYNTHESIS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Mark M. Friedman c/o Anthony Castorina

STREET: 20001 Jefferson Davis Highway, Suite 207

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk

COMPUTER: Twinhead, Slimnote 990TX

OPERATING SYSTEM: MS DOS version 6.2,

OPERATING SYSTEM: Windows version 3.11

SOFTWARE: Word for Windows version 2.0,

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/134.607A

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Friedman, Mark M.

REGISTRATION NUMBER: 33,883

REFERENCE/DOCKET NUMBER: 325/12

TELEPHONE: 972-3-5625553

TELEFAX: 972-3-5625554

TELEX:

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 498

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-134-607A-17

Query Match 7.5%; Score 152.5; DB 3; Length 498;

Best Local Similarity 22.1%; Pred. No. 1e-06;

Matches 71; Conservative 59; Mismatches 128; Indels 63; Gaps 16;

QY 5 YDLILVAGLANGLIALRLOQO--HPDMRILLIAGPEA--GGNHTWSPFEEDITLN-- 57

Db 82 FDVILIGAGPA---GLRLAEQSVKYGKVCVDSPLSMFNNYGVWVDFENIGLENC 137

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:26:38 ; Search time 45.165 Seconds
(without alignments)
2389.754 Million cell updates/sec

title: US-09-941-947A-30

effect score: 2021

sequence: 1 MQPHYDLILVAGLGLAFLA.....SGKPPVPVFALQAMTTHR 382

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

A Geneseq_29Jan04:*

1: Geneseq1900s:*

2: Geneseq1900s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Query	Score	Match	Length	DB ID	Description
1	2021	100.0	382	5	AAR22313	Pantoea s
2	2021	100.0	382	6	AAO16019	Pantoea s
3	2021	100.0	382	6	ABP96687	Pantoea s
4	1678	83.0	382	2	AAO7465	Polypepti
5	1678	83.0	382	2	AAW82258	C. utilis
6	1678	83.0	382	2	AAW87891	Protein e
7	1678	83.0	382	2	AAW99100	Erwinia u
8	1127	55.8	374	2	AAW01125	Lycopene
9	1127	55.8	374	2	AAW32474	Erwinia h
10	1126	55.7	374	2	AAW31987	Lycopene
11	1040.5	51.5	392	6	ABM70122	Phototrab
12	783.5	38.8	382	2	AAW06517	Flavobact
13	783.5	38.8	382	2	AAW69533	Flavobact
14	759.5	37.6	386	2	AAW79062	3 hydroxy
15	759.5	37.6	386	2	AAW87884	Protein e
16	759.5	37.6	386	2	AAW99096	Agrobacte
17	637	31.5	434	2	AAW95698	Erythroba
18	154.5	7.6	376	6	AAW31691	Rhodococ
19	154.5	7.6	498	3	AAW70397	Protein e
20	153.5	7.6	411	2	AAW76980	Lycopene
21	152.5	7.5	498	3	AAW70395	Protein e
22	152.5	7.5	498	3	AAW70395	Protein e
23	152	7.5	526	3	AAW54309	Amino aci
24	148.5	7.3	411	2	AAW00276	Cyanobact
25	148.5	7.3	498	2	AAW82242	Capsanthi

26	148.5	7.3	501	3	AAW54311	Amino aci
27	148.5	7.3	501	3	AAW54311	Arabidops
28	148.5	7.3	501	5	ABW2263	Herbicida
29	142.5	7.1	502	3	AAW54312	Amino aci
30	141.5	7.0	503	3	AAW54317	Amino aci
31	139.5	6.9	369	3	AAW54312	Arabidops
32	139.5	6.9	511	3	AAW90226	Marigold
33	139.5	6.9	511	3	AAW54316	Amino aci
34	139	6.9	498	2	AAW06454	Capsicum
35	139	6.9	498	3	AAW54313	Amino aci
36	139	6.9	500	3	AAW54315	Amino aci
37	136.5	6.8	517	4	AAW09798	Spinach 1
38	136	6.7	328	3	AAW54313	Arabidops
39	134.5	6.7	516	3	AAW90228	Marigold
40	134.5	6.7	516	3	AAW54310	Amino aci
41	133	6.6	407	6	ABU39747	Protein e
42	133	6.6	500	3	AAW54314	Amino aci
43	132.5	6.6	456	3	AAW54289	Consensus
44	131.5	6.5	524	5	ABG32343	Chimaeric
45	130.5	6.5	529	3	AAW54307	Amino aci

ALIGNMENTS

RESULT 1

AAE22313
ID AAE22313 standard; protein; 382 AA.

XX AC AAE22313;

XX AC AAE22313;

XX DT 25-JUL-2002 (first entry)

XX DT Pantoea stewartii lycopene cyclase (CrtY) enzyme.

XX DE Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

XX KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

XX KW aquaculture; enzyme; lycopene cyclase; CrtY.

XX OS Pantoea stewartii.

XX PN WO200218617-A2.

XX PD 07-MAR-2002.

XX PF 04-SEP-2001; 2001WO-US027420.

XX PR 01-SEP-2000; 2000US-0229858P.

XX PR 01-SEP-2000; 2000US-0229907P.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX PI Odom JW, Picataggio SK, Rouviere PE;

XX DR WPI; 2002-351711/38.

XX DR N-PSDB; AAD35511.

XX XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by

XX PT using microorganisms having a nucleic acid molecule encoding enzymes in

XX PT the carotenoid biosynthetic pathway and which metabolize single carbon

XX PS substrates.

XX PS Claim 21; Page 139-140; 156pp; English.

XX CC The invention relates to a method for producing carotenoid compounds. The

XX CC method comprises a transformed metabolising host cell, comprising

XX CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

XX CC encoding an enzyme in the carotenoid biosynthetic pathway, under the

XX CC control of regulatory sequences, and contacting the host cell with carbon

XX CC substrate to produce a carotenoid compound. The method is useful for

XX CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by

XX CC using microorganism having a nucleic acid molecule encoding enzymes in

CC the carotenoid biosynthetic pathway and which metabolise single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Pantoea stewartii lycopene cyclase (CrtY) enzyme used in the invention
 XX
 SQ Sequence 382 AA;

Query Match 100.0%; Score 2021; DB 5; Length 382;
 Best Local Similarity 100.0%; Pred. No. 8.1e-192;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOPHYDLILVAGLANGLIALRLOOQHDPMDILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 DB 1 MOPHYDLILVAGLANGLIALRLOOQHDPMDILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 QY 61 WIAPIVVRHWPDPYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLWHTAVSAVHAE 120
 DB 61 WIAPIVVRHWPDPYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLWHTAVSAVHAE 120
 QY 121 SVQLADGRIIHASTVIDGRGTYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDDQ 180
 DB 121 SVQLADGRIIHASTVIDGRGTYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDDQ 180
 QY 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 DB 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 QY 241 PITLTGDNRFQWQQPQACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVFTSSVHQTI 300
 DB 241 PITLTGDNRFQWQQPQACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVFTSSVHQTI 300
 QY 301 AHFAQORWQQQGFFFMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 DB 301 AHFAQORWQQQGFFFMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 QY 361 ILSGKPPVPVFAALQAIMTTHR 382
 DB 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 2
 AAO16019
 ID AAO16019 standard; protein; 382 AA.
 XX
 AC AAO16019;

DT 20-FEB-2003 (first entry)
 XX Pantoea stewartii lycopene Beta-cyclase.

XX Carotenoid; crt.

XX Pantoea stewartii.

XX WO200273995-A2.

XX 10-OCT-2002.

XX 25-JAN-2002; 2002WO-US002124.

XX 26-JAN-2001; 2001US-0264329P.

XX 04-MAY-2001; 2001US-028984P.

XX (CRGI) CARGILL INC.

XX De Souza ML, Kollmann SR, May CA, Schroeder WA;

XX WPI: 2003-075455/07.

XX N-PSDB; ABT14191.

XX Novel isolated nucleic acid useful e.g. to engineer host cells with the

PT ability to produce particular carotenoids and polypeptides useful in cell
 PT -free systems to make particular carotenoids.

XX Claim 14; Page 60-61; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of
 CC carotenoid (crt)-related proteins. The crt-related DNA and protein
 CC sequences of the invention are useful for engineering cells which are
 CC able to produce carotenoids. The present amino acid sequence represents a
 CC crt-related protein of the invention

XX SQ Sequence 382 AA;

Query Match 100.0%; Score 2021; DB 6; Length 382;
 Best Local Similarity 100.0%; Pred. No. 8.1e-192;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPHYDLILVAGLANGLIALRLOOQHDPMDILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 DB 1 MOPHYDLILVAGLANGLIALRLOOQHDPMDILLIEAGPEAGGNETWSFHEEDLTNQHR 60
 QY 61 WIAPIVVRHWPDPYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLWHTAVSAVHAE 120
 DB 61 WIAPIVVRHWPDPYQVRFPQRRRHVNSGYCVTSRHFAGILRQFGQHLWHTAVSAVHAE 120
 QY 121 SVQLADGRIIHASTVIDGRGTYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDDQ 180
 DB 121 SVQLADGRIIHASTVIDGRGTYTPDSALRVGFQAFICQEWQLSAPHGLSSPIIMDATVDDQ 180
 QY 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 DB 181 NGYRFVYTLPLSATALLEDTHYIDKANLQAEARQNIQDYAARQGWPLQTLREEQCAL 240
 QY 241 PITLTGDNRFQWQQPQACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVFTSSVHQTI 300
 DB 241 PITLTGDNRFQWQQPQACSGLRAGLFHTPTTGYSLPLAVALADRLSALDVFTSSVHQTI 300
 QY 301 AHFAQORWQQQGFFFMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 DB 301 AHFAQORWQQQGFFFMLNMLFLAGPAESRWVRVMQRFYGLPEDLIARFYAGKLTVTDLRL 360
 QY 361 ILSGKPPVPVFAALQAIMTTHR 382
 DB 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 3
 ABP96687
 ID ABP96687 standard; protein; 382 AA.
 XX
 AC ABP96687;

DT 03-JUN-2003 (first entry)

XX Pantoea stewartii lycopene cyclase SEQ ID NO:6.

XX Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
 KW crtI; crtB; crtZ; lycopene cyclase; enzyme; phytoene; carotenoid.

XX Pantoea stewartii.

XX WO2003016503-A2.

XX 27-FEB-2003.

XX 15-AUG-2002; 2002WO-US026647.

XX 15-AUG-2001; 2001US-0312646P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

HMG-CoA: 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtY: *crtY*; kw

Pantoea ananatis

KW carotenoid.
XX Pichia jadinii.
XX JP10248575-A.
XX 22-SEP-1998.
XX 12-MAR-1997; 97JP-00058012.
XX 12-MAR-1997; 97JP-00058012.
XX (KIRI) KIRIN BREWERY KK.
XX WPI; 1998-560727/48.
XX N-PSDB; AAV73182.
XX Gene useful for increase in carotenoid production - and preparation of carotenoid.
XX Example 2; Fig 15-17; 54pp; Japanese.
XX This invention describes a novel method for the preparation of carotenoids using genes and proteins isolated from Candida utilis. The invention specifically describes the isolation of a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence represents the Candida utilis crtY protein which is used in the method of the invention. (Updated on 17-OCT-2003 to standardise OS field)
XX Sequence 382 AA;
SQ
Query Match 83.0%; Score 1678; DB 2; Length 382;
Best Local Similarity 82.2%; Pred. No. 1e-157;
Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;
QY 1 MOPHYDLILVAGLANGIALRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEEDLTNQR 60
DB 1 MOPHYDLILVAGLANGIALRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEEDLTNQR 60
QY 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFQGOHLWHTAVSAVHAE 120
DB 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFQGOHLWHTAVSAVHAE 120
QY 121 SVQLADGRIIHAFTVIDRGYTPDSALVGVGFQAFIGQEWQLSAPHLSSPIIMDATVDOQ 180
DB 121 SVRLKGGVIGARAVIDGRGYAANSALSQVGFQAFIGQEWQLSAPHLSSPIIMDATVDOQ 180
QY 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIDYAAQCGWPLQTLREEOGAL 240
DB 181 NGYRFVYSLPSPTLLIEDTHYIDNATLDPECARQNICDYAAQCGWPLQTLREEOGAL 240
QY 241 PITLTGDRRQFWQOQPCACGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHCTI 300
DB 241 PITLSGNDAFWQRFACGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHCTI 300
QY 301 AHFAQRWQOQGFPRMLNRMFLAGPADSRVRVQRFYGLPEDIARFYAGKLTITDRLR 360
DB 301 THFAERWQOQGFPRMLNRMFLAGPADSRVRVQRFYGLPEDIARFYAGKLTITDRLR 360
QY 361 ILSGKPPVPVFAALQAIMTTHR 382
DB 361 ILSGKPPVPVFAALQAIMTTHR 382
RESULT 6
ID AAW87891
XX AAW87891 standard; protein; 382 AA.
XX AAW87891;
XX 17-OCT-2003 (revised)
DT 10-MAR-1999 (first entry)
XX

DE Protein encoded by the carotenoid biosynthesis gene crtY.
XX Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene; crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside; carotenoid diglucoside; crtY gene; crtZ gene; crtX gene; crtW gene; food additive.
XX Pantoea ananatis.
XX JP10327865-A.
XX 15-DEC-1998.
XX 29-MAY-1997; 97JP-00140460.
XX 29-MAY-1997; 97JP-00140460.
XX (KIRI) KIRIN BREWERY KK.
XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX WPI; 1999-099030/09.
XX N-PSDB; AAV84083.
XX New carotenoid glucoside(s) - used as food additives.
XX Disclosure; Page 22-23; 26pp; Japanese.
XX The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid diglucoside, in which all, or part of, carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtZ, crtX or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003 to standardise OS field)
XX Sequence 382 AA;
SQ
Query Match 83.0%; Score 1678; DB 2; Length 382;
Best Local Similarity 82.2%; Pred. No. 1e-157;
Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;
QY 1 MOPHYDLILVAGLANGIALRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEEDLTNQR 60
DB 1 MOPHYDLILVAGLANGIALRLOOQHPDMEILLIIBAGPEAGGNHTWSFHEEDLTNQR 60
QY 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFQGOHLWHTAVSAVHAE 120
DB 61 WTAPLVVHHWPDYQVRFPQRRHVNNSGYCVTSRHFAGILRQFQGOHLWHTAVSAVHAE 120
QY 121 SVQLADGRIIHAFTVIDRGYTPDSALVGVGFQAFIGQEWQLSAPHLSSPIIMDATVDOQ 180
DB 121 SVRLKGGVIGARAVIDGRGYAANSALSQVGFQAFIGQEWQLSAPHLSSPIIMDATVDOQ 180
QY 181 NGYRFVYTLPLSATALLIEDTHYIDKANLQAEARQNIIDYAAQCGWPLQTLREEOGAL 240
DB 181 NGYRFVYSLPSPTLLIEDTHYIDNATLDPECARQNICDYAAQCGWPLQTLREEOGAL 240
QY 241 PITLTGDRRQFWQOQPCACGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHCTI 300
DB 241 PITLSGNDAFWQRFACGLRAGLFHPTTGYSLPLAVALADRLSALDVFTSSVHCTI 300
QY 301 AHFAQRWQOQGFPRMLNRMFLAGPADSRVRVQRFYGLPEDIARFYAGKLTITDRLR 360
DB 301 THFAERWQOQGFPRMLNRMFLAGPADSRVRVQRFYGLPEDIARFYAGKLTITDRLR 360
QY 361 ILSGKPPVPVFAALQAIMTTHR 382
DB 361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 7

AW99100
AAW99100 standard; protein; 382 AA.

AAW99100;

17-OCT-2003 (revised)

14-MAY-1999 (first entry)

Erwinia uredovora crty protein sequence.

Beta-carotene hydroxylase; crty; crtB; crtE; crtI; xanthophyll;

metabolite.

Pantoea ananatis.

JP11046770-A.

23-FEB-1999.

07-AUG-1997; 97JP-00213648.

07-AUG-1997; 97JP-00213648.

(KIRI) KIRIN BREWERY KK.

WPI; 1999-208113/18.

N-PSDB; AAW19120.

Beta-carotin hydroxylase - useful for preparation of xanthophylls and their metabolites.

Disclosure; Page 15-17; 17pp; Japanese.

The present invention describes beta-carotin hydroxylase. Beta-carotene can be used in the preparation of xanthophylls and their metabolites. The present sequence represents an Erwinia uredovora crty protein sequence from the present invention. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 382 AA;

Query Match 83.0%; Score 1678; DB 2; Length 382;

Best Local Similarity 82.2%; Pred. No. 1e-157;

Matches 314; Conservative 31; Mismatches 37; Indels 0; Gaps 0;

1 M QHYDILVGLAGLIALRLQQOHPDMRILLIETAGPEAGNHTWSPFHEEDLTNQRH 60
1 M QHYDILVGLAGLIALRLQQOHPDMRILLIETAGPEAGNHTWSPFHEEDLTNQRH 60
61 WIAPLVHHWPDYQVRFPPQRRRHVNSGYCVTSRHPAGILRQOFGQHLMLHTAVSAVRAE 120
61 WIAPLVHHWPDYQVRFPPQRRRHVNSGYCVTSRHPAGILRQOFGQHLMLHTAVSAVRAE 120
121 SVOLADGRIHASTVIDGRGYPDSALRVGFOAFICQEWQLSAPHGLSPIMDATVDOO 180
121 SVOLADGRIHASTVIDGRGYPDSALRVGFOAFICQEWQLSAPHGLSPIMDATVDOO 180
121 SVRLKGVGIGARVAVIDGRGYAANSALSQVGFQAEWRLSHPHGLSPIMDATVDOO 180
181 NGYRFVYVTLPLSATALLIETHYIDKANIQAERARONIRDYAARQGWPLQTLREEQCAL 240
181 NGYRFVYVTLPLSATALLIETHYIDKANIQAERARONIRDYAARQGWPLQTLREEQCAL 240
241 PITLTGDNQFQWQOQOACSLRAGLFHTTGYSLPLAVALADRLSALDVFTSSVHOTI 300
241 PITLTGDNQFQWQOQOACSLRAGLFHTTGYSLPLAVALADRLSALDVFTSSVHOTI 300
301 AHPAQORWQOQOQFPRMLNMLFLAGPAERWRVWORFYGLPDLIARFYAGKLTITDLRL 360
301 AHPAQORWQOQOQFPRMLNMLFLAGPAERWRVWORFYGLPDLIARFYAGKLTITDLRL 360
361 ILSGKPPVPVFAALQAIMTTHR 382
361 ILSGKPPVPVFAALQAIMTTHR 382

RESULT 8

AAW01125

ID AAW01125 standard; protein; 374 AA.

XX AAW01125;

16-OCT-2003 (revised)

25-MAR-2003 (revised)

11-DEC-1996 (first entry)

Lycopene cyclase.

GQPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;

phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;

pigment; food colourant; chloroplast transit peptide; increase yield;

tobacco ribulose bis-phosphate carboxylase-oxygenase.

Pantoea agglomerans.

US5530188-A.

25-JUN-1996.

21-JUL-1993; 93US-00095726.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

30-OCT-1991; 91US-00785566.

(STAD) AMOCO CORP.

Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;

Ausich RL;

WPI; 1996-308823/31.

N-PSDB; AAT40795, AAT40796.

Increasing prodn. of total carotenoid(s) in a higher plant - by

transforming with vector encoding chloroplast transit peptide operably

linked to the Erwinia herbicola lycopene cyclase structural gene.

Example 17; Col 101-104; 99pp; English.

Manipulation of the DNA encoding the present sequence, lycopene cyclase,

by in frame linkage to the chloroplast transit peptide (AAW01124) of the

tobacco ribulose bis-phosphate carboxylase-oxygenase gene can lead to

increased production of total carotenoids in the chloroplast of

transformed plants as compared to native, non-transformed plants of the

same type. Beta-carotene is an effective and apparently harmless food

colourant and is also in the pathway for biological synthesis of further

C40 carotenoids such as zeaxanthin and zeaxanthin diglucoside. Other

enzymes involved in the carotenoid biosynthesis pathway include

geranylgeranyl pyrophosphate (AAW01119), phytoene synthase (W01121) and

phytoene dehydrogenase-4H (AAW01123). (Updated on 25-MAR-2003 to correct

PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 374 AA;

Query Match 55.8%; Score 1127; DB 2; Length 374;

Best Local Similarity 58.4%; Pred. No. 5.6e-103;

Matches 213; Conservative 53; Mismatches 99; Indels 0; Gaps 0;

6 DLIVGGLAGLIALRLQQOHPDMRILLIETAGPEAGNHTWSPFHEEDLTNQRHIAPL 65

3 DLIVGGLAGLIALRLQQOHPDMRILLIETAGPEAGNHTWSPFHEEDLTNQRHIAPL 62

66 VVHHWPDYQVRFPPQRRRHVNSGYCVTSRHPAGILRQOFGQHLMLHTAVSAVRAE 125

63 VAHAWPGYEVQPPDLRRRLRGYISITSERFABALHQALGENIWLNCVSSEVLPNSVRLA 122

Db 181 TGYRIYVLPSSRLLIEDTHYVDRGPPDVALSQATIAEYAKKHGKLGKLIRESGCL 240

QY 241 PITLTGDRQFQQ-QPQACSLRAGLFHTPTGYSLFLAVALADRLSALDVFTSSVHQT 299

Db 241 PITLTGDRQFQQ-QPQACSLRAGLFHTPTGYSLFLAVALADRLSALDVFTSSVHQT 300

QY 300 IAHFAQQRWQOQGFRLNRLFLAGPAESWRVMQRFYGLPEDLIARFVAGKLTVDRL 359

Db 301 LRDYARQOQWQRFRLNRLFLAGDQQRWQRFYGLPEDLIARFVAGKLTVDRL 360

QY 360 RILSGKPPVPVFAALQALMTTHR 382

Db 361 RILSGKPPVPVFAALQALMTTHR 383

RESULT 12

ID AAW06517 standard; protein; 382 AA.

AC AAW06517;

XX 17-OCT-2003 (revised)

DT 08-MAR-1997 (first entry)

DE Flavobacterium lycopene cyclase.

KW Carotenoid; lycopene; beta-carotene; echinenone; canthaxanthin;

KW zeaxanthin; adonixanthin; astaxanthin; crtY; lycopene synthase.

OS Flavobacterium sp. ATCC 21588; WT (ATCC 21588).

PN EP747483-A2.

XX 11-DEC-1996.

PD 29-MAY-1996; 9SEP-00108556.

PF 09-JUN-1995; 9SEP-00108888.

PR (HOFF) HOFFMANN LA ROCHE & CO AG F.

PA Hohmann H, Pasamontes L, Tessier M, Van Loon A;

PI WPI; 1997-023160/93.

DR N-PSDB; AAW45143.

XX Flavobacterium gene sequences encoding carotenoid biosynthesis enzymes - for the production of carotenoid(s), useful in foods and animal feeds.

PT Example 2; Fig 7; 80pp; English.

PS Lycopene synthase (AAW06517) is the product of the crtY gene identified in a carotenoid gene cluster (see also AAW45143) of Flavobacterium sp. R1534. This 42368 Da protein introduces the beta-ionone rings at both sides of lycopene to obtain beta- carotene. Enzymes of the Flavobacterium carotenoid biosynthetic pathway (see also AAW06515-18 and AAW0871) can be expressed, optionally with Alcaligenes beta-carotene beta-4-oxigenase, in host cells for the prodn. of lycopene, astaxanthin, beta-carotene, echinenone, zeaxanthin, canthaxanthin and adonixanthin. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 382 AA;

CC Query Match 38.8%; Score 783.5; DB 2; Length 382;

CC Best Local Similarity 43.7%; Pred. No. 7.9e-69;

CC Matches 165; Conservative 64; Mismatches 136; Indels 13; Gaps 5;

QY 5 YDILLVAGLGLIALRLQOQPMRILLIAGPAGNHTWTFHEEDLTINQHWIA- 63

Db 3 HDLLIAGLGLIALAVDRDPDARIWLDARGSPSQHTWSCHTDLS---PEMLAR 59

QY 64 --PLVGHHPDVQRFQRRHVNSSYCYCTSEHAGILRQGFQHLWHTAVSAVHBS 121

Db 60 LSPIRRGEWTQDEVAFPDHSRRLTTGYGSIAGALIGLLQ---GVDLRWNTHVATLDDTG 116

QY 122 VOLADGRIIHAFTVIDRGYTPDSALBVGCAPIGQWQLSAPHLSSPLIMDATVQQN 181

Db 117 ATLTDGSIIEAACVIDARGAVETPHLTVGFKFVGVEIETDAPHGVERPMINDATVQMD 176

QY 182 GYRFVYVTLPLSATALIEDTHVIDKANLQAFRAQNIQDYAARQGWPLQTLREOQALP 241

Db 177 GYRFYLLFPFSPTRILLIEDTRYSDGDLDDGALAAQASLDVAAREGWTGQE-MRERGILP 235

QY 242 IELTGDNRQFQQOQPOAC--SGLRAGLFHTPTGYSLFLAVALADRLSALDVFTSSVHQT 299

Db 236 IALAHDALIGFWRDHAQGAVPVGLGAGLFHPVTGYSLFLAVALADRLSALDVFTSSVHQT 294

QY 300 IAHFAQQRWQOQGFRLNRLFLAGPAESWRVMQRFYGLPEDLIARFVAGKLTVDRL 359

Db 295 VRGWALDRADRDRLNRLFLAGCPDPRYRLQRFYRLPQLIERFYAGRLTLADRL 354

QY 360 RILSGKPPVPVFAALQAI 377

Db 355 RIVTGRPPILSQAVRCL 372

RESULT 13

ID AAW69533 standard; protein; 382 AA.

XX AAW69533;

AC 10-AUG-1999 (first entry)

DT Flavobacterium sp. R1534 crtY gene product lycopene cyclase.

DE Carotenoid; pigment; canthaxanthin; R1534; crtB; prephytoene synthase; crtI; phytoene desaturase; crtY; lycopene cyclase; GGPP synthase; crtB; crtW2396; beta-carotene beta-oxygenase; food product; fermentation.

OS Flavobacterium sp.

XX JP10155497-A.

PN 16-JUN-1998.

PD 02-DEC-1997; 97JP-00348653.

PF 02-DEC-1996; 96EP-00810839.

PR (HOFF) HOFFMANN LA ROCHE & CO AG F.

PA WPI; 1998-391048/34.

DR N-PSDB; AAW40146.

XX Preparation of carotenoid - comprises fermentation with transformed cell.

XX Claim 1; Fig 25; 80pp; Japanese.

CC The invention describes the preparation of carotenoid pigments e.g. canthaxanthins using a cell transformed by a vector having DNA sequences (a) to (e) or substantially homologous sequences. (a) a DNA sequence (crtB) coding GGPP synthase of Flavobacterium sp. R1534; (b) a DNA sequence (crtB) coding prephytoene synthase of Flavobacterium sp. R1534; (c) a DNA sequence (crtI) coding phytoene desaturase of Flavobacterium sp. R1534; (d) a DNA sequence (crtY) coding lycopene cyclase of Flavobacterium sp. R1534; and (e) a DNA sequence (crtW2396) coding beta-carotene beta-oxygenase of a microbe E-396 (FERM SP-4283). The carotenoid or a carotenoid mixture can also be used in preparation of food products. The method is an improved method of fermentation for carotenoid production

XX Sequence 382 AA;

CC Query Match 38.8%; Score 783.5; DB 2; Length 382;

CC Best Local Similarity 43.7%; Pred. No. 7.9e-69;

PS Disclosure; Page 12-13; 26pp; Japanese.
XX
CC The present sequence represents a protein involved in carotenoid
CC biosynthesis. The specification describes astaxanthin diglucosides and
CC adonixanthin-3'-glucosides. The specification also describes a method for
CC the preparation of a carotenoid glycoside, in which all, or part of,
CC carotenoid biosynthesis genes crtB, crtI, crtY, crtZ, crtX or crtW
CC are introduced to a microbe or plant and expressed. The transformed
CC organism is cultured and astaxanthin diglucosides, adonixanthin-3'-
CC glucosides, and/or astaxanthin monoglucoside are collected. The
CC carotenoid glucosides are used as food additives
XX
SQ Sequence 386 AA;
Query Match 37.6%; Score 759.5; DB 2; Length 386;
Best Local Similarity 43.0%; Pred. No. 2e-66;
Matches 164; Conservative 65; Mismatches 137; Indels 15; Gaps 8;
2Y 5 YDLILVAGLANGLIALFLQQHDPDRILLIE--AGPEAGNHTWSPHEEDLTINQHRWI 62
Db 3 HDVLLAGLANGLIALRAARPDRLVLLDHAAGPSDG--HTWSCHDFDLSPD---WL 57
2Y 63 A---PLVVHHPDYQVRPQRRRHVNGYYCVTSRHPFAGILRQOFGOHLWLHTAVSAVHA 119
Db 58 ARLLKPLREANWPDQVRPRPHARELATYGSGLDGAALADAVVRSAGAIRW-DSDIALLDA 116
2Y 120 ESVQLADGRIIHASTVIDGCGYTPDSALRYGQFAFIGEWQLSAPHGLSSPIIMDATVDQ 179
Db 117 QGATLSCGTRIEAGAVLDGRGQPSRHLTVGFQKVGVEIETDRPHGVPRPNIMDAIVTQ 176
2Y 180 QNGYRFVYTLPLSATALIEDTHYIDKANQAERARQNIIDYAAQGMPLQTLLEEQGA 219
Db 177 QDGYRFIYLLPFSPTRIILEDTRYSDGDLDDALAAASHDYARQQGW-TGAEVRRERGI 235
2Y 240 LPITLTGNRQFWQQP--QACSGIRAGLPHPTTGYSLPLAVALADRLSALD-VFTSSSV 296
Db 236 LPIALAHDAAGFWADHAAGPVFVGLRAGFFHPTVGTYSCLPYAAQVADVAGLSPGTDAL 295
2Y 297 EQTIAHFAQQWQQGFFRMLNRLMLFLAGPAESRWVMQRFYGLPEDLIARFYAGKLTVT 356
Db 296 RGAIRDYVIDRARDRFLRLNRLMFRGCAPDRRYTLQRFYRMPHGLIERFYAGRLSVA 355
2Y 357 DLRLILSGKPPVPVPAALQAI 377
Db 356 DQLRIVTGKPPILGTAIRCL 376

Search completed: February 29, 2004, 14:44:00
Job time : 48.265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:33:49 ; Search time 35.4397 Seconds
(without alignments)
3837.172 Million cell updates/sec

file: US-09-941-947A-28

effect score: 2231
sequence: 1 MSFVAVIAPPPFHVRLQN.....EQAMRTQPVLSGGDYATAT 431

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Maximum Match 100%

Listing first 45 summaries

atabase : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2231	100.0	431	2 Q8GCS3	Q8GCS3 pantoea ste
2	1794	80.4	431	2 Q47843	Q47843 pantoea agg
3	1579.5	70.8	431	2 Q8VUJ7	Q8VUJ7 pantoea agg
4	558.5	25.0	419	16 P72650	P72650 synchocyst
5	262	11.7	407	16 Q9S1V2	Q9S1V2 streptomyc
6	251.5	11.3	397	16 Q825U1	Q825U1 streptomyc
7	247	11.1	412	16 Q9S0P6	Q9S0P6 streptomyc
8	243.5	10.9	389	2 Q83X61	Q83X61 streptomyc
9	239.5	10.7	461	2 Q9XC67	Q9XC67 streptomyc
10	238.5	10.7	402	16 Q81E86	Q81E86 bacillus ce
11	234.5	10.5	402	16 Q81R64	Q81R64 bacillus an
12	228.5	10.2	389	16 Q8DSH9	Q8DSH9 streptococ
13	219.5	9.8	379	2 Q93H13	Q93H13 streptomyc
14	219	9.8	417	2 Q86304	Q86304 streptomyc
15	217	9.7	424	2 Q68841	Q68841 streptomyc
16	213	9.5	446	2 Q83WG5	Q83WG5 streptomyc

17	210.5	9.4	427	16	Q98EL9	Q98e19 rhizobium
18	207	9.3	397	16	Q81CG5	Q81cg5 bacillus ce
19	207	9.3	406	16	Q97TQ3	Q97tq3 clostridium
20	204	9.1	392	16	Q34539	Q34539 bacillus su
21	201	9.0	397	2	Q8KNE0	Q8kne0 micromonosp
22	198.5	8.9	418	16	Q9RYI3	Q9ryi3 deinococcus
23	197.5	8.9	418	16	Q9ADH3	Q9adh3 streptomyc
24	197	8.8	450	16	Q89RS0	Q89rs0 bradyrhizob
25	195.5	8.8	287	16	Q31853	Q31853 bacillus su
26	187.5	8.4	421	2	Q8KHE4	Q8khe4 nocardia ae
27	181.5	8.1	540	5	Q9VCL3	Q9vcl3 drosophila
28	178.5	8.0	402	2	Q8KHU5	Q8khu5 actinosynne
29	178.5	8.0	492	10	Q9SBL1	Q9sbl1 sorghum bic
30	177	7.9	594	5	Q16243	Q16243 caenorhabdi
31	175.5	7.9	414	2	Q98595	Q98595 nocardia ae
32	175	7.8	407	16	Q97FM0	Q97fm0 clostridium
33	175	7.8	418	16	Q8PJG6	Q8pjg6 xanthomonas
34	172.5	7.7	530	11	Q9R110	Q9r110 cavia porce
35	172	7.7	388	16	Q7Y01	Q7y01 mycobacteri
36	172	7.7	420	16	Q33282	Q33282 mycobacteri
37	171	7.7	382	2	Q9F2F9	Q9f2f9 streptomyc
38	170.5	7.6	401	2	Q8KNC3	Q8knc3 micromonosp
39	169.5	7.6	333	16	Q81CW9	Q81cw9 bacillus ce
40	167	7.5	422	2	Q70023	Q70023 streptomyc
41	164.5	7.4	388	2	Q9PEA1	Q9pea1 streptomyc
42	164.5	7.4	406	2	Q93MW2	Q93mw2 nocardia br
43	164.5	7.4	541	11	Q91WS7	Q91ws7 mus musculu
44	164	7.4	392	16	Q49841	Q49841 mycobacteri
45	163	7.3	529	11	Q8VIF8	Q8vif8 cavia porce

ALIGNMENTS

RESULT 1

ID Q8GCS3 PRELIMINARY; PRT; 431 AA.
 AC Q8GCS3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Zeaxanthin glucosyl transferase.
 GN CCTX.
 OS Pantoea stewartii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=66269;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 8200;
 RA deSouza M.L., Kollmann S.R., Schroeder W.A.;
 RT "Carotenoid Biosynthesis (WO 02/079395 A2).";
 RL Submitted (OCT-2002) to the EMBL/GenBank/DDAJ databases.
 DR EMBL; A1166713; AAN85597.1; --
 DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR006326; UDPGT_MGT.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT_1.
 DR TIGRFAMs; TIGR01426; MGT; 1.
 KW Transferase.
 SQ SEQUENCE 431 AA; 46611 MW; A9BC082567039732 CRC64;

Query Match 100.0%; Score 2231; DB 2; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.2e-159;
 Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVAVIAPPPFHVRLQNLAQELVARGHVRVTPPQQHDCALVTGSDIGFQTGLQTHP 60
 |||||
 Db 1 MSFVAVIAPPPFHVRLQNLAQELVARGHVRVTPPQQHDCALVTGSDIGFQTGLQTHP 60
 |||||
 QY 61 PGSLSHLHLAAHPGPSMLRLINEMWNTSDMLCRLPAFAHALQIEGVIVDQMEPAGAV 120

```
1b 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120
1y 121 VASASGLPFFVSACALPLNREPGLPLAVNPFYGTSDAAREYTTSEKIYDMLMRHDRV 180
1b 121 VASASGLPFFVSACALPLNREPGLPLAVNPFYGTSDAAREYTTSEKIYDMLMRHDRV 180
1y 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDFPRKALPDCCHAVGLPQOQTPGSS 240
1b 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDFPRKALPDCCHAVGLPQOQTPGSS 240
1y 241 TSYPSPDKPRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
1b 241 TSYPSPDKPRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
1y 301 GDIQVDFADQSAALSQAOLITTHGGMNTVLDIAASRTPLLAFLAFDQPGVASRIVVHG 360
1b 301 GDIQVDFADQSAALSQAOLITTHGGMNTVLDIAASRTPLLAFLAFDQPGVASRIVVHG 360
1y 361 IGKRASRTTSHALAQIRSLTNTDYPQRMTKIQALRLAGGTPAAADIIVEQAMRTCP 420
1b 361 IGKRASRTTSHALAQIRSLTNTDYPQRMTKIQALRLAGGTPAAADIIVEQAMRTCP 420
1y 421 VLSGQDYATAL 431
1b 421 VLSGQDYATAL 431

RESULT 2
347843
ID Q47843 PRELIMINARY; PRT; 431 AA.
AC Q47843
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CrtX.
EN CrtX.
RN CrtX.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=94236237; PubMed=8180698;
RX To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
RA Chang Y.S., Liu S.T.;
RT "Analysis of the gene cluster encoding carotenoid biosynthesis in
RT Erwinia herbicola Ehol3.";
RL Microbiology 140:331-339(1994).
DR EMBL; M90698; AAA21261.2; -.
DR PIR; S52583;
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR00213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 431 AA; 47070 MW; ED08B121D77C19CE CRC64;

Query Match 80.4%; Score 1794; DB 2; Length 431;
Best Local Similarity 82.4%; Pred. No. 2.5e-126;
Matches 355; Conservative 19; Mismatches 57; Indels 0; Gaps 0;

QY 1 MSHFAVIAPPFSSHVRALONLAQELVARGHRTVFPQCHCKALVTGSDIGFQTVGLQTHP 60
Db 1 MSHFAIAAPPFSSHVRALONLAQELVARGHRTVFPQCHCKALVTGSDIGFQTHPHIGHVAGVEESH 60

QY 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120
Db 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAL 120

QY 121 VASASGLPFFVSACALPLNREPGLPLAVNPFYGTSDAAREYTTSEKIYDMLMRHDRV 180
Db 121 VASASGLPFFVSACALPLNREPGLPLAVNPFYGTSDAAREYTTSEKIYDMLMRHDRV 180
```

```
QY 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDFPRKALPDCCHAVGLPQOQTPGSS 240
Db 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDFPRKALPDCCHAVGLPQOQTPGSS 240
QY 241 TSYPSPDKPRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Db 241 TSYPSPDKPRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
QY 301 GDIQVDFADQSAALSQAOLITTHGGMNTVLDIAASRTPLLAFLAFDQPGVASRIVVHG 360
Db 301 GDIQVDFADQSAALSQAOLITTHGGMNTVLDIAASRTPLLAFLAFDQPGVASRIVVHG 360
QY 361 IGKRASRTTSHALAQIRSLTNTDYPQRMTKIQALRLAGGTPAAADIIVEQAMRTCP 420
Db 361 IGKRASRTTSHALAQIRSLTNTDYPQRMTKIQALRLAGGTPAAADIIVEQAMRTCP 420
QY 421 VLSGQDYATAL 431
Db 421 VLSGQDYATAL 431

RESULT 3
Q8VUU7
ID Q8VUU7 PRELIMINARY; PRT; 431 AA.
AC Q8VUU7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CrtX protein.
EN CrtX.
RN CrtX.
OS Pantoea agglomerans pv. milletiae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=182454;
[1]
RN SEQUENCE FROM N.A.
RA Kamiunten H. Hirata R.;
RT "Isolation and characterization of carotenoid biosynthesis genes from
RT Pantoea agglomerans pv. milletiae Wist 801.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB076662; BAB79601.1; -.
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR00213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
SQ SEQUENCE 431 AA; 47439 MW; F7E124F2632A8EF7 CRC64;

Query Match 70.8%; Score 1579.5; DB 2; Length 431;
Best Local Similarity 70.9%; Pred. No. 3e-110;
Matches 304; Conservative 53; Mismatches 71; Indels 1; Gaps 1;

QY 1 MSHFAVIAPPFSSHVRALONLAQELVARGHRTVFPQCHCKALVTGSDIGFQTVGLQTHP 60
Db 1 MSHFAVIAPPFSSHVRALONLAQELVARGHRTVFPQCHCKALVTGSDIGFQTVGLQTHP 60

QY 61 PGSLSHLHLAAHPPLGSPMLRINEMARTSDMLCRLPAAFHALQIEGVIVDQMEPAGAV 120
Db 61 AGSLAHTLQLAHPPLGSPMLKLINEAMARTSDMLCRLPVLSTALDGVIVDQMEPAGAL 120

QY 121 VASASGLPFFVSACALPLNREPGLPLAVNPFYGTSDAAREYTTSEKIYDMLMRHDRV 180
Db 121 AAEALNPFYVSACALPLNREPGLPLAVNPFYGTSDAAREYTTSEKIYDMLMRHDRV 180

QY 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDFPRKALPDCCHAVGLPQOQTPGSS 240
Db 181 IAHACRMGLAPREKLHCFSPPLAQISOLIPELDFPRKALPDCCHAVGLPQOQTPGSS 240
QY 241 TSYPSPDKPRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 299
Db 241 QPRIFPHGDRPRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
```

```

300 GGDGIQVDFADQSAALSQAQLTIHTGGKNTVLDIAISRTPELLALPIAFDPQPGVASRIYVH 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 ASHVQVDFADQQAALAQAQLVITGGKNTVLDGINHLTPELLTIPIAFDPQPGVAARVVH 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 GICKRASRTTTHSHALQIRSLITNDYPQMTKIOAASLAGGTAAADIIVEQAMRTCO 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 GIGRASRTTTHSHMARQLQTLADESVQCKMTIRSLAQAGTTLAADIIVEQMLTRQ 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 PVLSGDQYA 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 PVLTRHYA 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

P726S0 PRELIMINARY; PRT; 419 AA.

P726S0;
01-FEB-1997 (T-EMBLrel. 02, Created)
01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
Zeaxanthin glucosyl transferase.
CTX OR SLR125.
Synecocystis sp. (strain PCC 6803).
Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
NCBI_TaxID=1148;
[1]
SEQUENCE FROM N.A.
MEDLINE=97061201; PubMed=8905231;
Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
Miyaajima N., Hiroseawa M., Sugliura M., Sasamoto S., Kimura T.,
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
Shimpo S., Takeuchi C., Wada T., Watanabe A., Yanada M., Yasuda M.,
Tabata S.;
"Sequence analysis of the genome of the unicellular cyanobacterium
Synecocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
DNA Res. 3:109-136(1996).
EMBL; D90899; BAA16652.1; -.
PIR; S74500; S74500.
GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
GO; GO:0016740; P:transferase activity; IEA.
GO; GO:0008152; P:metabolism; IEA.
InterPro; IPR006326; UDPGT_MGT.
InterPro; IPR02213; UDP_Gluco_trans.
Pfam; PF00201; UDPGT_1.
TIGRFAMs; TIGR01426; MGT; 1.
Transferase; Complete proteome.
SEQUENCE 419 AA; 45330 MW; BAE7DFDE2592CB3 CRC64;

Query Match 25.0%; Score 558.5; DB 16; Length 419;
Best Local Similarity 31.7%; Pred No. 1,1e-33;
Matches 132; Conservative 81; Mismatches 196; Indels 7; Gaps 5

1 MSFPAVIAPFFSHVFRALQNLAQELVARGHRVTFPQQHDCALVTGSDIGQFQTVGLQTHP 60
1 MTHPGLLCPATTGHLNTMLPLGKELQQRGHTVTMFEGVLDAQAKTLAGLNFQAIAITTEPP 60
61 PGSLSHLLHLAHPGLPSMLR-LINEMARTSDMLCRELPAAFAHALQIEGVIVQMEPAGA 119
61 LGAQEFNABLKGSLGKIALQYTVAKITQKAAAFEEAPGVMAKAGVEVLINVOVSEGG 120
120 VVAEASGLPVSVCALPLNRRPGLPVLAVMPFXYGTSDAABERYTTSEKIVYDMLMRHR 179
121 TIGRLGIPFISCSAVLNREPTTPVATPYPDPSWLGQRNRLGVLGNRAATKPITA 180
180 VIAHHCNMGALPREKLHCFSPLAQISOLIPELDFPRKALPDCFHAVGLRQFQGTGPGS 239
181 LINDYRQWNLPAQSPNDRISPLAQISQDPAAPFPRECLPSRPHFTGTFHNSVGRDIA 240
240 STSYFPSPDFPRIPASLGTQLQGHRYLFTTIKACEEVDQAQLLHACGGLSATQAQELAR 299
241 DFPWEQITDQPIVASLGTIONLQMTFKIIAEACWLDQAQLIIS-LGGAKLSEMPALP- 298

```

Qy	300	GDIOQVDFPDQSAALSQAQLTTHGKNVTLDIAIRSTRPTLLALPLAFDOPGVASRLVYH	357
Db	299	-GNPLWYNAPQLSLLQRTALTATTHAGINTTLECLANNAVPMVAIPANOPGVAAARIAWA	357
Qy	360	GGKG--RASRFTTSHALARQIRSLTNTDTPQRMTKIQALRLAGGTTPAAADIVEQ	413
Db	358	GVGBEPIPLSKLNTNN-LRAALEKVLTEDSYKRNLTQLQQAIKTAGGLTKAADIIIEQ	412

RESULT 5			
Q981V2	PRELIMINARY;	PRT;	407 AA.
AC	Q981V2;		
DT	01-NAV-2000 (TREMBLrel. 13, Created)		
DT	01-NAV-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DT	Putative glycosyl transferase.		
GN	SCO0040 OR SCJ4.21.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=1902;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Saunders D.C., Harris D.;		
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;		
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
EX	MEDLINE=97000351; PubMed=8843436;		
RA	Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,		
RT	Kinashi H., Hopwood D.A.;		
RT	"A set of ordered cosmids and a detailed genetic and physical map for		
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";		
RT	Mol. Microbiol. 21:77-96(1996).		
RL	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2) / M145;		
EX	MEDLINE=21996410; PubMed=12000953;		
RA	Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,		
RA	Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,		
RA	Harper D., Baceam A., Brown S., Chandra G., Chen C.W., Collins M.,		
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,		
RA	Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,		
RA	Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,		
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,		
RA	Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,		
RA	Hopwood D.A.;		
RT	"Complete genome sequence of the model actinomycete Streptomyces		
RT	coelicolor A3(2).";		
RL	Nature 417:141-147(2002).		
DR	EMBL; AL939104; CABS2955.1; -.		
DR	PIR; T37104; T37104.		
DR	CO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.		
DR	CO; GO:0016740; P:transferase activity; IEA.		
DR	CO; GO:0008152; P:metabolism; IEA.		
DR	InterPro; IPR006326; UDPGT_MGT.		
DR	InterPro; IPR002213; UDP_Gluco_trans.		
DR	Pfam; PF00201; UDPGT; 1.		
DR	TIGRFAMS; TIGR01426; MGT; 1.		
KW	Transferase; Complete proteome.		
SO	SEQUENCE 407 AA; 43498 MW; A61E94CF608B84A0 CRC64;		

Query Match 11.7%; Score 262; DB 16; Length 407;
Best Local Similarity 26.8%; Pred. No. 1.8e-11;
Matches 133; Conservative 59; Mismatches 165; Indels 112; Gaps 18;

2Y 1 MSFAVIAPPPFHVSRALQNLQALVARGHYTF--EQHCDCKALVTGSDIGFQTVGLQT 58
1 MSTIAFLNIGHGHIPTLPVAAELVRGHTVTHYTPAFREIATGANV-----RL 53
59 HPGSLSHLLHAAHP--GPSMLRLINEMARTSMKRELPAAPHALQIEG---VIVDQ 113
54 YPGGD-----QPLDPDPAPITIMEALARTSLDL---LPAVLADLRDORPOLIIVHDS 101
114 MEPAGVAEASGLPFVSVACALPLNREPCPLVAVMPPEYGTSDAARERTTSEKIYDML 173
102 ACWGAALAAVGLPVPVSTFYNNRVHVPSP-----TRASRE----- 139
174 MRHRRVIAHACRMGLAPREKLUHCFSPALQISQL-----IPELDFPRKALPDCPHAVGP 229
140 -----LLAGAAAR-----PRLAGYVGAARLARRRFAATGVPLVD-----LAD 177
230 LROFGTGGSTSVFSP-----PEKPR-----IPASLGTLQ 261
178 IRQPLNLVTSRAFOVAEVEFDRSYRVPVGSIGARPDPSFPVNLRDPLVYASLGTVEN 237
262 HRYGLPFTIAKACEEVDQAQLLHACGGLSATQAGELARGDQVDFADQSAALSQAQLT 321
238 ADPLLLTFATLSPL-AGVVVSTGCTPPAALGELP--GNVLAERSVPQLEVLDRALF 294
322 ITGCGMNTVDALASRTPLALPLATDQPCVASRIWVGIGK--RASFT--TSHALAQ 377
295 ITGCGMNSVNEALFAGVPLLLVPGQADQWVAERVVGLGAGLSITENTEDSVRAVAR 354
378 IRSLLTNDYFQMTKIQALRLAGTTPAAADIVEQMR 416
355 ---LLEDSRYRAATSGLRATQHEAGGYRRAADELEGYLR 390
RESULT 6
Q825U1
ID Q825U1 PRELIMINARY; PRT: 397 AA.
AC Q825U1;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative glycosyl transferase.
GN SAV7358.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12592562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP005050; BAC75069.1; --
DR GO: GO:0016740; F:transferase activity; IEA.
DR Interpro: IPR006326; ODPRT_MGT.
DR TIGRFAWS: TIGR01426; MGT; I.
KW Transferase; Complete proteome.
SQ SEQUENCE 397 AA; 40868 MW; 77065B1CC04D6FE7 CRC64;

Query Match 11.3%; Score 251.5; DB 16; Length 397;
Best Local Similarity 26.6%; Pred. No. 1.1e-10;
Matches 115; Conservative 61; Mismatches 186; Indels 71; Gaps 18;
QY 1 MSFAVIAPPPFHVSRALQNLQALVARGHYTF--EQHCDCKALVTGSDIGFQTVGLQTHP 60
DB 1 MGRFVFPVPLVGVNPAVGTAAALAAARGHDIWAGHPVLVRLGAGADAVVFPFAL---P 57
QY 61 PGSLSHLLHAAHPGLSPMLRLINE--MARTSMKRELPAAPHALQIEGVIVDQMEPAG 118
DB 58 EQLSRPAGLK-----GPAFOFLWESFLVPLADAMAPGVRAAIEAYDPVVCDCQAVAG 113
QY 119 AVVASASGLPFVSVACALPLNREPCPLVAVMPPEYGTSDAARERTTSEKIYDMLRRHD 178
DB 114 ALVASESLGRPVATGATT---SAELVDPLAGMP-----KVAANL-----D 149
QY 179 RVIAHACRM---GLA--PREKLUHCFSPALQISQLIPELDFPRKALPDCPHAVGP--LR 231
DB 150 GLLGLRLRRITGGAGADPR-----FSPGVLAITYTTRALLGPVE-LPDRVVLVGPSVAA 202
QY 232 PQQTPGSGSTSVFSPDKPRIFASLGTQGHRYGLFRTIAKACEEVDQAQLLHACGGLS- 290
DB 203 RPAGDDDPFEMWLEASALPTVLSLGTANNDAGARFLNAA-----ASAL-----GGIAD 251
QY 291 ATOAGELARGGDIO-----VDFADQSAALSQAQLTITHCGMNTVDALASRTPLLA 342
DB 252 RVRAVLVDFGVVHPVDPDTVLVRYVPLALLERLDVAVVCHAGNTVCEALHGVPLVV 311
QY 343 LPLAPDQGVASRIWVGIGKASRTTSHA--LAROIRSLTNDYFQMTK-IQALR 399
DB 312 AFIEDDQPIVAQVVDAGAGVRL-RFGRAAARIGAAVEAVLPDAQHQRKAAEAVGESF 370
QY 400 LAGGTTPAAADIVE 412
DB 371 RAGGSESAADRL 383
RESULT 7
Q850P6
ID Q850P6 PRELIMINARY; PRT: 412 AA.
AC Q850P6;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Glycosyl transferase.
GN AVEHI OR SAV945.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RT anthelmintic macroide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

381 -LLNTDYPQMTKIOAALRAGGTAAADIVEBOAM 415
353 ELTADEVRSDALRAVRAEGTTRAAADIIIESAL 388

RESULT 9

99XC67
ID Q9XC67 PRELIMINARY; PRT; 461 AA.
AC Q9XC67;
XT 01-NOV-1999 (T-EMBLrel. 12, Created)
YT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
JT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
JE Mycosyl transferase TyICV.
KN TyICV.
XS Streptomyces fradiae.
XC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
XC Streptomycineae; Streptomycetaceae; Streptomyces.
XN NCBI_TaxID=1906;
[1]
SEQUENCE FROM N.A.
XP STRAIN-T59235;
XQ MEDLINE=50121747; PubMed=10658660;
XA Bate N., Butler A.R., Smith I.P., Cundliffe E.;
XT "The mycosyl-biosynthetic genes of Streptomyces fradiae, producer of
tylosin.";
XJ Microbiology 146:139-146 (2000).
XK EMBL; AF147704; A041824.1; .
XL GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
XM GO; GO:0016740; P:transferase activity; IEA.
XN GO; GO:0005975; P:carbohydrate metabolism; IEA.
XO GO; GO:0030259; P:lipid glycosylation; IEA.
XP InterPro; IPR004276; Glyco_trans_28.
XQ InterPro; IPR006326; UDPGT_MGT.
XR Pfam; PF03033; Glyco_transf_28; 1.
XS TIGRFAMs; TIGR01426; MGT; 1.
XT Transferase.
XU SEQUENCE 461 AA; 50120 MW; 4A61AE1F06A2F59 CRC64;

Query Match 10.7%; Score 239.5; DB 2; Length 461;

Best Local Similarity 23.3%; Pred. No. 1.1e-09;
Matches 108; Conservative 58; Mismatches 179; Indels 119; Gaps 16;
QY 1 MSHFAVIAEPFFSHVRALONLAQELVARGHRTVFQQHDC--KALVTGSDIGFQTVGLQT 58
DB 58 MAHIAFIFLPAAGHVNPTLGVAELAAARHRTYALPEDMADRAVRVGA----- 106
QY 59 HPPGSLHLLHAAHPLGPMRLINEMARTSDMLCRE----- 96
DB 107 -----RAVTPLDREFR-----ADMVPEESDEYTDGEFLKVLWLLDTT 148
QY 97 ---LPAAFHALQIEGVIQDMPA---GAVVAEASGLPFVSVACALPLNREPGLPLAVM 149
DB 149 ADTLPLESAFADRDVDVANDPSTFTGLLAGKWDIPVIRSTPVSASHNHALH--P 205
QY 150 PFYGTSD---AARERTYTSKIDMLMRH---DRVIAHACRMG-----LAPR----- 193
DB 206 PFEGGAQVDPALIELTARAER---LLKEHSTSDPVAFAATVQSGPLFYKPRYFOYA 261
QY 194 ----EKLHCFSPLAQISQLIFELDFPRKALPDCPHAVPLRQPGTSGSSTSYFSPDK 249
DB 262 GETFDNRHVEGCA-----PRAS-----PH-----GTWQRPEDGR 292
QY 250 PRIFASLGLTQGHRYGLFTTIKACEEVDQAQLLAHOGGLSATQAGELARGDQIQQVDF 309
DB 293 PLWVSLGTYINERPGIFRACVAFRDRPWNILLVGLGGLGAGDLGPLPE--NVLVRDFV 350
QY 310 DQSAALSQAQLTTHGGMNTVLDASTRTPILALPLAFDQPGVASRIYVHGIGK-RASRF 368
DB 351 PLGCVLPHRTDLLVNGGTSTANEALAKGVPIVAMPEPEPATARRIAELDLDGWLPG 410
QY 369 TTSHALARQIRSLNTDYPQMTKIOAALRAGGTAAADIVE 412
DB 411 VTAEKLSGAQRVLTDDRIRKGLDRMRGEIRRAGGPAVADVE 454

RESULT 10

Q81EB6
ID Q81EB6 PRELIMINARY; PRT; 402 AA.
AC Q81EB6;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Macrolide glycosyltransferase (EC 2.4.1.-).
GN BC2066
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=22608415; PubMed=12721630;
RX Ivanova N., Sorokin A., Anderson I., Galleron N., Gandelon B.,
RA Kapatral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin I., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
Bacillus anthracis.";
RL Nature 423:87-91 (2003).
RM EMBL; AE017004; AAP09035.1; .
DR GO; GO:0016758; P:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR006326; UDPGT_MGT.
DR InterPro; IPR002213; UDP_Gluco_trans.
DR Pfam; PF00201; UDPGT_1.
DR TIGRFAMs; TIGR01426; MGT; 1.
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 402 AA; 45743 MW; 75B9D25EBC38D82 CRC64;

Query Match 10.7%; Score 238.5; DB 16; Length 402;

Best Local Similarity 21.6%; Pred. No. 1e-09;
Matches 92; Conservative 81; Mismatches 199; Indels 53; Gaps 10;
QY 1 MSHFAVIAEPFFSHVRALONLAQELVARGHRTVFQQHDCALVTGSDIGFQTVGLQTHP 60
DB 1 MANVLVINFPGEGHINPTLAVSELIQRGETVVSVCIEDYRKVKVEATGAEFVFP----- 54
QY 61 PGSLHLLHAAHPLGPMRLINEMARTSDMLCRSLPAAFHALQIEGVIQDMPAGAV 120
DB 55 ENFLSQINIMERNVEGSPPLTSLHMEASERIVTOIVETKEEKYDYLINHPFVGRI 114
QY 121 VAEASGLPFVSVACALPLNR-----EPGLPLAVMPPEYGTSDAARERTYS-----EK 168
DB 115 IVNITQLPSVSSCTTFAVNOYINFHDEQSRQVDEMPLYQSCLAGMERWNRKQGMKNS 174
QY 169 IYDMLMRHDEVTAAHACRMGLAPRKLHCHCPSPLAQISQLIFELDFPRKALPDCPHAVG 228
DB 175 MYDLMNHGPDITVI-----TSKE-----YQPRSEV-----FDESXKFGV 209
QY 229 PLRQPGTSGSSTSYFSPD---KPRIFASLGLTQGHRYGLFTTIKACEEVDQAQLLAH 285
DB 210 PSIAITRKEVGS---PFTEDLKNKVFIFSMGTVPFNEQPALYKCFEAFKVDATVVLVW 265
QY 286 CGGLSATQAGELARGDQIQQVDFADQSAALSQLFTTHGGMNTVLDASTRTPILALPL 345
DB 266 GKKINISQFENIPEK--NFKLYNTVPQLEVLQHADVFTVHGGMNSSEALYGVPLVPIPV 323
QY 346 AFDQPGVASRIYVHGIGKRASR-FTTSHALARQIRSLNTDYPQMTKIOAALRAGGT 404
DB 324 TGDQPFVAKSLTGVAGITILNRNELTSELLRETVKVMDVTFKENSILKVGESLRNAGY 383
QY 405 PAAD 409
DB 384 QRAVE 388

RESULT 11

Db 378 RNAGGYKRAVD 388

RESULT 12

Q8DSH9 PRELIMINARY; PRT; 389 AA.

AC Q8DSH9; 01-WAR-2003 (TrEMBLrel. 23, Created)

DC 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative glycosyltransferase.

GN SMU.1806.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [!]

RP SEQUENCE FROM N.A.

RC STRAIN=UAI59 / ATCC 700610 / Serotype C;

RX MEDLINE=22395063; PubMed=12397186;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,

RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).

RL EMBL; AE015008; AAN59431.1; -.

DR GO: GO:0005489; F:electron transporter activity; IEA.

DR GO: GO:0016758; F:transferase activity, transferring hexosyl . . ; IEA.

DR GO: GO:0016740; F:transferase activity; IEA.

DR GO: GO:0006118; P:electron transport; IEA.

DR InterPro: IPR000345; CytC_heme_B8.

DR InterPro: IPR006326; UDPGT_MGT.

DR InterPro: IPR002213; UDP_Gluco_trans.

DR Pfam: PF00201; UDPGT; 1.

DR TIGRFAMs: TIGR01426; MGT; 1.

DR PROSITE: PS00190; CYTOCHROME_C; 1.

DR PROSITE: PS00375; UDPGT; 1.

KW Transferase; Complete proteome.

SW Query Match 10.2%; Score 228.5; DB 16; Length 389;

SW Best Local Similarity 23.4%; Pred. No. 5.6e-09;

SW Matches 103; Conservative 66; Mismatches 190; Indels 81; Gaps 17

QY 2 SHEFAVAP--PFFSHVVALQVLAQELVARGHRTFFQHDCKALVTGSDIGFTVGLQTH 59

DB 5 THYNTIMLVNLPFSGHTNPTGLGAILVOLGHNTVYNAPDWQSKIERTGAHPV---YDR 61

QY 60 PGSLSHLLHLAAHPLGLPSMLRLINEMARTSDMLCRLPAAFHALQIEG-----VIVDM 114

DB 62 YDLSLSE-----RQKKVKSWSAAAYQTVLVRMGKDYDCLIVEML 98

QY 115 EPAGAVVAEASGLPFSVACALPLNRBPGLPLAVMFEYGTSDAARERTTSEKIYDWM 174

DB 99 FEPFGKALADRLNPAYRLPFSYSLNEK-----ILTFGKTGGF---YLTS--IFRFTF 146

QY 175 RHEDRVIAHACMLAPREKLHCF---SPLA-QISQLIPDELFPF---RXALPDC--- 223

DB 147 LRH-----LLSR-KLQTFQLAYDCLKRISLMTPELNTFYTVREFOIDADTF 193

QY 224 ----FHAVGP-LRQGGTGGSTSYFSPDKPRIPASLGLQCHRGCLERTAKACEVD 278

DB 194 DENHYQYVGPNSINRPVEPFPDFTPF-----KNPIIYISLGLTLNRSVSPFKCKIKAFENP 249

QY 279 AQLLHAHCGELSATQAGELARGGDIQVDFADQSAALSQLATITHGCGNTVLDAISRT 338

DB 250 YGIIISLGNRIKSGQLGMP--ANVHLYSVFPQQLILERASLFLTHGGWNSVNEALVYGC 307

QY 339 PLLALPLAFDPGVASRIYVHGIGKGRASRTTS-HALARQIRSLLTNTDYPQBWTKIA 397

DB 308 PMLVTPVGNDDPQVRAQVQVADILHGCKLCKLNPLAQEIKQAAHTILKXSSYKNTIWHFQKI 367

398 LRLAGTTPAAADIVEQAMRT 417
368 AQTAGNTUJAQIINDLQI 387

RESULT 13

933H13
ID Q93H13 PRELIMINARY; PRT; 379 AA.
AC 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 25, Last annotation update)
DE Putative glycosyl transferase.
GN Streptomyces avermitilis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21477403; PubMed=11572948;
RX Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
"Genome sequence of an industrial microorganism Streptomyces
avermitilis: Deducing the ability of producing secondary
metabolites";
ZL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMBL; AB070941; BAE69204.1; -;
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR006326; UDPGT_MGT.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT_1.
DR TIGRFAMs; TIGR01426; MGT; 1.
DR Transferase.
KW TRANSFERASE.
SQ SEQUENCE 379 AA; 38968 MW; B5FED865CB11D5C CRC64;

Query Match 9.8%; Score 219.5; DB 2; Length 379;
Best Local Similarity 26.5%; Pred. No. 2.6e-08;
Matches 109; Conservative 60; Mismatches 172; Indels 71; Gaps 18;
QY 22 AQELVARGHRTVFQHQDCKALVTGSDIGFQVGLQTHPPGSLSHLLHLAAHP;GPSMLR 81
DB 4 AAALARGHDIAGHPLELVRLGLAGADAVVFPAL---PEDGLSRPAGLK---GPAAFQ 56
QY 82 LINE--MARTSDMLCRLPAFHALQIEGVIVDQMERAGAVVAESGLPFFSVACALPLN 139
DB 57 FLWSEPLAVLADAMPQVRAATEAYDPDVVVCDOQAVAGALVAESLGRPWTSATT---S 113
QY 140 REPGLPLAVWPPEYGTSDAAREYTTSEKIYDMLRHRDRIAHACRM---GLA-PRE 194
DB 114 AELVDPLAGMP-----KVAAML-----DGLGLRLRRTGGAGAADPR- 151
QY 195 KLHCFSPALAQISQLIPELDFRKALPDCFHAGVP--LRPQGTGSGSTSYFPSPDKPRI 252
DB 152 -----PSPGVLYATTTRALLGPVE-LPDRVLMVGPSVAARPAQDPDFPWEWLEASALPTV 205
QY 253 PASLGTQCHRYGLERTIAKACEVDQAQLLHACGSL-ATQAGELARGGDIQ----- 304
DB 206 LWSLGTANDAGARFLNAR-----AGAL-----GGIADRVAVLVDPGGWHEPVPDVT 254
QY 305 -VYDFADQSAALSQAQLTTHGGMNTVLDAISKRTPLALPLAPDQPGVASRIYTHGIGK 363
DB 255 LVRRYVFPQALLERLDVAVVCHAGHNTVCEALWHGVLPVAPIRDDQPIVAQVVDAGAV 314
QY 364 RASRETTTSHA--LARGIRSLTNTDYPORYTK-IQAALLRAGGTPAAADIVE 412
DB 315 RL-RFGRAAARIGAARVAVLDPQAQCHRAARAEAVGESFRAAGSGESAARLIE 365

RESULT 14

086304
ID 086304 PRELIMINARY; PRT; 417 AA.
AC 086304;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Macrolide glycosyl transferase.
GN MGT.
OS Streptomyces ambofaciens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1889;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC23877;
RX MEDLINE=99445176; PubMed=10517588;
RA Pernodet J.L., Gourmelin A., Blondelet-Rouault M.H., Cundliffe E.;
"Dispensable ribosomal resistance to spiramycin conferred by srmA in
the spiramycin producer Streptomyces ambofaciens";
ZL Microbiology 145:2355-2364(1999).
DR EMBL; AJ223970; CHA11707.1; -;
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR006326; UDPGT_MGT.
DR InterPro; IPR002213; UDP_gluco_trans.
DR TIGRFAMs; TIGR01426; MGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
DR Transferase.
KW TRANSFERASE.
SQ SEQUENCE 417 AA; 45427 MW; B693EF18F6FCASFD CRC64;
Query Match 9.8%; Score 219; DB 2; Length 417;
Best Local Similarity 24.8%; Pred. No. 3.2e-08;
Matches 111; Conservative 68; Mismatches 185; Indels 84; Gaps 17;
QY 2 SHPAVIAPPFFSHVRLQNLQAELVARGHRTVF-----PQHQDCKALVTGSDIGFQVGL 56
DB 22 AIIAMFSTRAHGHVNPVSLVIRLVARGHRTVTAIPPLFAE---KVAETGAEFKLWNSTL 78
QY 57 QTHPPGSLSHLLHLAAHP--LGPSMLRLINEMARTSDMLCRLPAFHALQIEGVIVDQ 114
DB 79 ----RGP-----DADEPDAGTTPLDNVPEPLDDAIQALPQLIAAYEGDEPDLVLDIT 127
QY 115 EPAGAVVAESGLPFFSVACAL---PLNREPLPLAVMPF--EYGTSDAAREYTTSP- 167
DB 128 SYPARVLAAHKGVPANVSLSNLVAMGSEYEEVGRPTWEEPLKTERGRAYDARFGLMKEN 187
QY 168 ---KIYDMLRHRDRIAHACRMGLAPREKLHCFSPALAQISQLIPELDFPR 217
DB 188 GUTEDPDPFFVGRDRLSLVLPKALQPHADRV-----DEKTHTFVGACQGDRA----- 234
QY 218 KALPDCPHAVGPLRQPGTSGSTSYFPSPDKPRIPASLGTQCHRYGLERTIAKACEV 277
DB 235 -----ABGDWRRPESA-----EKVVLVLSIGSFTTKRPAPFACVAFGAL 274
QY 278 DAQULLAHQCGLSATQAGELARGGDIQVYDFADQSAALSQAQLTTHGGMNTVLDAIAS 336
DB 275 PGWHVVLQVGRHVDPAELGDVPE--NVEVRSVWPQLAILKQADLFVTHAGAGCSQGLAT 332
QY 337 RPLPALLPLAPDQPGVASRIYTHGIGKRSRPTTSHALAQIR-----SLTWTYDPOMT 392
DB 333 ATPIVAVQAVDQFGNADMLQGLGVGR---HLPTERTABALRAAGLALVEDPFAVRUK 389
393 KIQAALRAGGTPAAADIVE---QANRT 417
QY
DB 390 EIQAGMAREGGTRRADLTAEALAAAT 417
RESULT 15
086841
ID 086841 PRELIMINARY; PRT; 424 AA.
AC 086841;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)

Job time : 40.4397 secs

```
I 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
R Glycosyltransferase OleI.
R OLEI.
S Streptomyces antibioticus.
C Bacteria; Actinobacteridae; Actinomycetales;
C Streptomycineae; Streptomycetaceae; Streptomyces.
K NCBI_TaxID=1890;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 11891;
X MEDLINE=98343801; PubMed=9880207;
A Quiros L.M., Aguirrezabalaga I., Olano C., Mendez C., Salas J.A.;
L Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
R EMBL; AF055579; AAC12648.1; -.
R PTR; T51110; T51110.
R GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
R GO; GO:0016740; F:transferase activity; IEA.
R GO; GO:0008152; P:metabolism; IEA.
R InterPro; IPR006326; UDPGT_MGT.
R InterPro; IPR002213; UDP_gluco_trans.
R Pfam; PF00201; UDPGT; 1.
R TIGRFAMs; TIGR01426; MGT; 1.
R PROSITE; PS00375; UDPGT; 1.
W transferase.
Q SEQUENCE 424 AA; 45384 MW; FE7F90DB9A1C11EC CRC64;

Query Match 9.7%; Score 217; DB 2; Length 424;
Best Local Similarity 24.7%; Pred. No. 4.6e-08;
Matches 111; Conservative 69; Mismatches 185; Indels 84; Gaps 19;

Y 3 HFAVIAPPFFSHVRLQNLQAEVLVARGHRTFFQHQDKALV-----TCSDIGFQTVGLQ 57
b 14 HIGFFNIPGHVNSLIGVIGLVARGQVSIGITDFGCAQVKGAGRAVAVYGF----- 67
Y 58 THPPGSLHLHLAHPGLGPMRLINEMARTSDMLC---RELPAAFHALQIEG---VIV 111
b 68 -----ILPEEFNPEELAEHQSRWACSLAEAFVLPQLRTATPTTGRDLIVY 115
Y 112 DQMEPAGAVVAESGLPFV-----SVACA-----LPLNREPGLPL---AVMPFEYGTSDA 158
b 116 DIASWPAPVLGRKWDIPFVQLSPSTSVAYEGFEDVPVQDPTADRGEEAAAP--AGTGDA 173
Y 159 ARERYTTSSKIYDMLMRHRDVIA---HH-----ACRMGLAPREKLHCFSPLAQISOLI 210
b 174 --BEGAEAE---DGLVRFTRLSAFLEEGVDTPATEFLIAPNRCIVGCRAP----SQIK 224
Y 211 PELDFPRKALPDCFHAVGPL---RQPGQT---PGSSTSYFPSPDKPRIPASIGTQGHRY 264
b 225 GD-----TVGDNVTFVGPTYGDRSHQGTWEGFGHG-----RPVLLIALGSAFTDEL 270
Y 265 GLPFTAKACEEVDQAQLLHAHCGLSLNTAQAGELARGDILQWDFADQSAALSOAQLTIH 324
b 271 DFYRTCLSAVDGDLWHVLSVGRFVDPADLGEVP--PNVEVHOWVPQDLILTRASAFITH 328
Y 325 GGNNTVLDAISRFTLLALPLAFDPQGVASRIYVHGIGKRASR-FTTSHALARQIRSLIT 383
b 329 AGKGTMEALSNAVPMVAVPQIAEQTMNAERIIVELGLGRHIPDQVTAETKLEAVLAVAS 388
Y 384 NTDPQRTKIQAAALRLAGCTPAAADIVE 412
b 389 DPGVAERLAARQREIRAGGARAADILE 417
```

Search completed: February 29, 2004, 14:50:54

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

on on: February 29, 2004, 14:27:19 ; Search time 7.26404 Seconds

(without alignments)

3089.496 Million cell updates/sec

file: US-09-941-947A-28

sequence: 1 MSHPFVIAFPFPHSHVRLQN.....EQAMTCQPVLSGQDYATATL 431

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Maximum Match 0%

Listing first 45 summaries

database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1790	80.2	431	1	CRTX PANAN
2	1018.5	45.7	413	1	CRTX_EWHE
3	228.5	10.2	430	1	OLSD_STRAT
4	183	8.2	502	1	UDCL_RABIT
5	181.5	8.1	418	1	MGT_STRLI
6	174.5	7.8	541	1	CGT_HUMAN
7	167	7.5	455	1	UPOG_HORVU
8	164.5	7.4	541	1	CGT_MOUSE
9	164.5	7.4	541	1	CGT_RAT
10	162	7.3	535	1	UD11_RAT
11	158	7.1	414	1	YF24_MYCTU
12	156	7.0	530	1	UD88_RAT
13	154.5	6.9	527	1	UD11_HUMAN
14	150.5	6.7	505	1	UGTF_CABEL
15	150.5	6.7	530	1	UD19_HUMAN
16	150	6.7	528	1	UD12_MOUSE
17	149	6.7	533	1	UD12_MOUSE
18	148.5	6.7	471	1	UPO2_MAIZE
19	147.5	6.6	530	1	UD1A_HUMAN
20	145.5	6.5	471	1	UPO3_MAIZE
21	145.5	6.5	530	1	UD12_HUMAN
22	143.5	6.4	471	1	UPO1_MAIZE
23	142	6.4	529	1	UD8S_HUMAN
24	140.5	6.3	530	1	UD17_HUMAN
25	139.5	6.3	379	1	Y336_METWA
26	139.5	6.3	527	1	UD11_RAT
27	138.5	6.2	530	1	UD8F_HUMAN
28	138	6.2	533	1	UD12_RAT
29	137	6.1	533	1	UD11_HUMAN
30	134	6.0	531	1	UD13_RAT
31	134	6.0	531	1	UD15_RAT
32	133.5	6.0	361	1	MURG_CAUCR
33	133.5	6.0	388	1	Y453_METAC

ALIGNMENTS

RESULT 1

ID	CRTX PANAN	STANDARD	PRT	431 AA
AC	P21686;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DB	Zeaxanthin glucosyl transferase (EC 2.4.1.-)			
GN	CRTX OR UGT101.			
OS	Pantoea ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoea.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
RX	MEDLINE=91072214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashina K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway			
RT	by functional analysis of gene products expressed in Escherichia			
RT	coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
CC	- FUNCTION: CATALYZES THE GLYCOSYLATION REACTION WHICH CONVERTS			
CC	ZEAXANTHIN TO ZEAXANTHIN-BETA-DIGLUCOSIDE.			
CC	- PATHWAY: Carotenoid biosynthesis.			
CC	- SIMILARITY: Belongs to the UDP-glycosyltransferase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; D90087; BAAL4125.1; -			
DR	PIR; B37802; B37802.			
DR	InterPro; IPR002213; UDP_gluco_trans.			
DR	Pfam; PF00201; UDPGT; 1.			
DR	PROSITE; PS00375; UDPGT; 1.			
RW	Transferase; Carotenoid biosynthesis.			
SK	SEQUENCE 431 AA; 47241 MW; 9CE7A798608BC9D CRC64;			

Query Match	80.2%;	Score 1790;	DB 1;	Length 431;
Best Local Similarity	80.5%;	Pred. No. 4.7e-131;		
Matches 347;	Conservative 29;	Mismatches 55;	Indels 0;	Gaps 0;
QY	1	MSHPFVIAFPFPHSHVRLQNLAQELVAERHRTVFQOQDCKALVTGSDIGFTVGLQTHP	60	
DB	1	MSHPFVIAFPFPHSHVRLQNLAQELVAERHRTVFQOQDCKALVTGSDIGFTVGLQTHP	60	
QY	61	PGSLSHLHLAHPGPSMLRLINEMATSNLRELPAAFALQIEGVIVDQEPAGAV	120	
DB	61	PGALTRVHLAHPGPSMLKLINEKATNTDMLCRLPQAFNDLAVDGVIVDQEPAGAL	120	

61 -----PGSLSHLLHLLAHLPLGSMRLINEMARTSDMLCRLPAAFH 102
77 FNSITSDAFASQKMRNIFSGRLTAI-----ELFDILDHYTKCDLM-----VGNH 121
103 ALQIEG-----VIVDMPAGAVAAEAGLFPVSACALPLNRPFG--LPLAVMPFE 152
122 AL-IOGLKKEKFDLLVDPNMOCGFVIAHLLGVKAVFSTGLWYPAEYGAAPLAYVP-E 179
153 YOTSDAARERTTSKIDYMLMRHRDVRVIAHACRMGLA-----PREKLHCFSPLAQISQ 208
180 FNSLLTDRMN-----LIQRMXNTGVVLI1SRLGVSTFLVLEK-----YERIMQKYN 223
209 LIPE-----LDPEKALPDCFHAGVLPQPGTSGSSTSYFSPD 248
224 LPKESMDLVHSGSLWMLCTDVALEFPRPLPNVYVGGILTKPAS-----PLPE 274
249 KPRI-FASLGTGLQGHRYGLFRTIAKACEEVDQAQLLHACGGLSATOAGELAR----- 299
275 DIQRWNGANEGHGFVLVSFGAGKVLSEDLANKL-----AGALGRLPQKVIWRF 323
300 -----GGDIQVDFADQSNALSOAQLT--ITHGGMNTVLDALASRTPLALLPLAFDQ 349
324 SGFKPKNLGNNTKLEWLPONDLLGHSKIKAFSLHGGLSIFETWYHGVVGVVGLPFGDH 383
350 PGVASRIVYHIGIKRAS-RFTTSHALARQIRSLTNTDYPQRMTKIQAAALAGGTP 405
384 YQTRVQAKGNGILLEWTKVTEKELYALVKVINNPYSYRQRAKLSIHKDQFCHP 440

RESULT 7

FOG HORVU STANDARD; PRT; 455 AA.
214726;
01-APR-1990 (Rel. 14, Created).
01-NOV-1997 (Rel. 35, Last sequence update).
28-FEB-2003 (Rel. 41, Last annotation update).
E Flavonol 3-O-glucosyltransferase (SC 2.4.1.91) (UDP-glucose flavonoid
3-O-glucosyltransferase) (Bronze-1).
N B21 OR UGT71A2.
S Hordeum vulgare (Barley).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
C Triticeae; Hordeum.
X NCBI_TaxID=4513;
[1]
P SEQUENCE FROM N.A.
N STRAIN=cy. Abyssinian 2231;
X MEDLINE=91329682; PubMed=2151660;
X Wise R.P., Rohde W., Salamini F.;
A "Nucleotide sequence of the Bronze-1 homologous gene from Hordeum
vulgare."
T Plant Mol. Biol. 14:277-279 (1990).
L C
C -I- FUNCTION: In the presence of other necessary color factors, this
C glycosylation reaction allows the accumulation of anthocyanin
C pigments.
C -I- CATALYTIC ACTIVITY: UDP-glucose + a flavonol = UDP + flavonol
C 3-O-D-glucoside.
C -I- PATHWAY: Anthocyanin biosynthesis.
C -I- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
C or send an email to license@sib-sib.ch).
C
C EMBL: X15694; CAA33729.1; --
R PIR: S14919; XUBHFC.
R InterPro: IPR002213; UDP_glucosyltrans.
R Pfam: PF00201; UDPGT; 1.
R PROSITE: PS00375; UDPGT; 1.

KW Transferase: Glycosyltransferase.
SQ SEQUENCE 455 AA; 47079 MW; PFD8BC2E2F103AC9C CRC64;

Query Match 7.5%; Score 167; DB 1; Length 455;
Best Local Similarity 24.1%; Pred. No. 1.6e-05;
Matches 112; Conservative 56; Mismatches 194; Indels 102; Gaps 21;

QY 3 HPAVLAPFPFHSVRALQNLQALQELVA--RGRVTFVFFQOHD-----CKALVTGSDIGFQTV 54
Db 7 HIAVAVFPFSSEAVALFSAFALAAAPAGTSLSLFTTTADNAAQLKAGALPGNLRFFVEV 66
QY 55 --GLQTHPPGSLSHL-----LHLAHPGLFPMRLINEMARTSDMLCRLPAAFAHALQ 105
Db 67 PDGV---PPGETSCISPPRRMDLFMAAAEAGGVRL-----EAAAC-----ASAGGAR 111
QY 106 IEGVTVDMQEPAGAVVAEASGLFPFVSV---ACALPLNREPGLFLAVMPFPGTSDAARE 161
Db 112 VSCVVDGAPVWT-ADAAAGAGAPWAVVTAASCAL-----LAHLRTDALR- 155
QY 162 RYTTSEKIDYMLMRHRDVRVIAHACRMGLAPRE-----KLHCFSPLA--QISQ-- 209
Db 156 ----RDVGDOAASRADELLVAHAGLGGYVRDLDPGVVSGDFNVYISLLVHRQARLPK 210
QY 210 -----IPELDPP-----RKALPDCFHAGVGLR-QPGTSGSSTSYFPS----- 246
Db 211 AATAVALNTFPGDLPDLIALAELPNCL-PLGPHLLPGAEPTADTNEAPADPHGCLA 269
QY 247 -----PDKEPRI-FASLGTGLQGHRYGLFRTIAKACEEVDQAQLLHACGGLSATOAGELARGG 301
Db 270 WLDRRPARSVAVVSGTNTATARPDELQELAAGLEASGAPFLMSLGVVAAAPRGFLERAP 329
QY 302 DIQVDFADQSNALSOAQLT--TITHGGMNTVLDALASRTPLALLPLAFDQGVASRIV-Y 358
Db 330 GL-WVEMAPQGVLRHAAVGAFFVTHAGNWSVMEGVSSGVPMACRFFPGDQTNARSVASV 388
QY 359 HGIGKRASTFTSHALARQIRSLTNTDYPQRMTKIQAAALRLAG 402
Db 389 WGTGTAFDGMTRGAVANAVATLLRGEDGERMKAKAQELQAMVG 432

RESULT 8
CGT_MOUSE
ID CGT_MOUSE STANDARD; PRT; 541 AA.
AC Q64576; Q61634;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-hydroxyacylsphingosine 1-beta-galactosyltransferase precursor
DE (SC 2.4.1.45) (UDP-galactose-ceramide galactosyltransferase) (Ceramide
DE UDP-galactosyltransferase) (Cerebroside synthase).
GN UGT8 OR CGT OR UGT4.
OS Mus musculus (Mouse).
OC Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Bosio A., Binczek E., Stoffel W.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC Coetzee T., Li X., Fujita N., Marcus J., Suzuki K., Francke U.,
RA Popko B.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: CATALYZES THE TRANSFER OF GALACTOSE TO CERAMIDE, A KEY
CC ENZYMIC STEP IN THE BIOSYNTHESIS OF GALACTOCEREBROSIDES, WHICH
CC ARE ABUNDANT SPHINGOLIPIDS OF THE MYELIN MEMBRANE OF THE CENTRAL
CC NERVOUS SYSTEM AND PERIPHERAL NERVOUS SYSTEM.
CC -I- CATALYTIC ACTIVITY: UDP-galactose + 2-(2-hydroxyacyl)sphingosine =
CC UDP + 1-(beta-D-galactosyl)-2-(2-hydroxyacyl)sphingosine.
CC -I- PATHWAY: Galactocerebroside biosynthesis.
CC -I- SIMILARITY: Belongs to the UDP-glycosyltransferase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X92122; CAAG3090.1; -
EMBL; X92123; CAAG3091.1; -
EMBL; X92124; CAAG3091.1; JOINED.
EMBL; X92125; CAAG3091.1; JOINED.
EMBL; X92126; CAAG3091.1; JOINED.
EMBL; X92177; CAAG3091.1; JOINED.
EMBL; U48896; AAC33576.1; -
EMBL; U48892; AAC33576.1; JOINED.
EMBL; U48893; AAC33576.1; JOINED.
EMBL; U48894; AAC33576.1; JOINED.
MGD; MGI:109522; Ygt8.
InterPro; IPR002213; UDP_gluco_trans.
Pfam; PF00201; UDPGT; 1.
PROSITE; PS00375; UDPGT; 1.
Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
Microsome.
CHAIN 1 20
SIGNAL 21 541
POTENTIAL.
2-HYDROXYACYLSPHINGOSINE 1-BETA-
GALACTOSYLTRANSFERASE.
TRANSMEM
472 492
CARBOHYD 78 78
N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 333 333
N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 442 442
N-LINKED (GLCNAC. . .) (POTENTIAL).
T -> S (IN REF. 2).
H -> R (IN REF. 2).
CONFLICT 458 458
CONFLICT 511 511
CONFLICT 524 524
CONFLICT 533 534
HI -> RV (IN REF. 2).
K -> R (IN REF. 2).
SEQUENCE 541 AA; 61137 MW; B76F80A9B5326E8 CRC64;

Query Match 7.4%; Score 164.5; DB 1; Length 541;
Best Local Similarity 20.3%; Pred. No. 3.1e-05;
Matches 93; Conservative 76; Mismatches 197; Indels 85; Gaps 17;

QY 6 VIAPP--FSSHVRAQLAQELVARGHRTVFQHQDCALVTGSDIG-PQTVGLQTHP-- 60
DB 24 IIVPTMFSSHLVYFKTLASALHERGH-----HTVLLSEGRDIAPSNFYSLQRYPGI 76
QY 61 -----PGSLSHLLLAHPPLGFSMLRNLINEMARTSDMLC--RELPA 100
DB 77 FNSTTSDAFLQSKMNFISGRITAV-----ELVDILHYTKNCDDMMVGNQALIQ 126
QY 101 FHAIQTEGVVDQMEPAGAVARASGLPVSACALPLNREPG--LPLAVMPFYGTSDA 158
DB 127 LKKEKFDLLVDPNOMCGFVIAHLGVKYAVFSTGLMYPAEVGPAPLAYVP-EFNSLLT 185
QY 159 ARERYTTSK-----IYDLMERHRRVIAHACRMGLAPREKLHCF--GPLAQ 205
DB 186 DRNFMFLRMKNTGVYLSRIGVSVFLVPKYERIMQ-----KYNLLPAKSMYDLVHGSLLWM 241
QY 206 ISQLPELDPKALPDCHAVGPIRQCGTGGSTSYFFSPDPKPRIPASLGLTQGHRYG 265
DB 242 LCTDV-ALSEFPRTPLNPNVYVGGILTKPAS-----PLPEDLRWVSGAQEHGVFLV 291
QY 266 LPRTIKACEEYDAQLLAHCCGL-----SATOAGELARGGDIQVDFADQSAALSQ 317
DB 292 SFAGVYLYSEDIANKLAGALGRLPQKVIWFSGTKPKNL--GNNTKLIBFLPQNDLLGH 349
QY 318 AQLT--ITHGGMNTVLDIAISRTPLALPLAPDPQGVASRIVTHGIGKGRASRTT-TSHAL 374
DB 350 SNIRAPLSGGLNSIFETMTHGVPPVVG-PLFGDHYDTMTVRQAKGMGILLEMNTVTEGSL 409
QY 375 ARQIRSLNTDYPQRMWKIQALRIAGTGP 405

DB 410 YDALVKVNNPSYQRAQKLSEHKQPGHP 440

RESULT 9
CGT RAT
ID CGT RAT STANDARD; PRT; 541 AA.
AC Q09426;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-Hydroxyacylsphingosine 1-beta-galactosyltransferase precursor
(EC 2.4.1.45) (UDP-galactose-ceramide galactosyltransferase) (Ceramide
UDP-galactosyltransferase) (Cerebroside synthase).
GN UGT8 OR CGT OR UGT4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WiStar; TISSUE=Brain;
RX MEDLINE=94052143; PubMed=7694285;
RT "Ceramide UDPgalactosyltransferase from myelinating rat brain:
purification, cloning, and expression";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10265-10269 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=94358923; PubMed=7521399;
RA Stahl N., Jurevics H., Morell P., Suzuki K., Popko B.;
RT "Isolation, characterization, and expression of cDNA clones that
encode rat UDP-galactose: ceramide galactosyltransferase";
RL J. Neurosci. Res. 38:234-242 (1994)
CC [-] FUNCTION: CATALYZES THE TRANSFER OF GALACTOSE TO CERAMIDE, A KEY
ENZYMATIC STEP IN THE BIOSYNTHESIS OF THE MYELIN MEMBRANE OF THE CENTRAL
ARE ABUNDANT SPHINGOLIPIDS OF THE MYELIN MEMBRANE OF THE CENTRAL
NERVOUS SYSTEM AND PERIPHERAL NERVOUS SYSTEM.
CC [-] CATALYTIC ACTIVITY: UDP-galactose + 2-(2-hydroxyacyl)sphingosine =
UDP + 1-(beta-D-galactosyl)-2-(2-hydroxyacyl)sphingosine.
CC [-] PATHWAY: Galactocerebroside biosynthesis.
CC [-] TISSUE SPECIFICITY: BRAIN, RESTRICTED TO THE OLIGODENDROCYTE-
CONTAINING CELL LAYERS OF CEREBRUM AND CEREBELLUM.
CC [-] SIMILARITY: Belongs to the UDP-glycosyltransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L21698; AAA16108.1; -
EMBL; U07683; AAA50212.1; -
PIR; A48801; A48801.
InterPro; IPR002213; UDP_gluco_trans.
Pfam; PF00201; UDPGT; 1.
PROSITE; PS00375; UDPGT; 1.
Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
Microsome.
CHAIN 1 20
SIGNAL 21 541
POTENTIAL.
2-HYDROXYACYLSPHINGOSINE 1-BETA-
GALACTOSYLTRANSFERASE.
TRANSMEM
472 492
CARBOHYD 78 78
N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 333 333
N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 442 442
N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 541 AA; 61126 MW; 260D7603170151BB CRC64;

Query Match 7.4%; Score 164.5; DB 1; Length 541;
Best Local Similarity 20.3%; Pred. No. 3.1e-05;

Matches 108, Conservative 56; Mismatches 173; Indels 108; Gaps 25;

QY 24 ELVARGHRTVFFQCHDCKALVTGSDIGP-QTVGLQTHPPGSL-----SHLHLAAHPL 75

DB 23 ELORGHGV-----CLA-VPNLIQFVETAGLSAVAVGSRDSQSQDLQFLHNAWKLO 74

QY 76 GPSMLRLINE-----WARTSDMLCRLPAPFAHA-LOIEGVIVDOMEPAQAVAEASG 126

DB 75 NP--IKLLREMAPVTEGNALSAMU---TPVAGADLLLTGQIQEVV---ANVAEHG 126

QY 127 LPFVSV-ACALPNRREPLAVMPFEGYSDAAREYTTSEKIYDMLRRHRDRIAHHA 185

DB 127 IPLAALHFYPRANGIAPP-ARLP-----APLVRSTITADMLY-WRMTK----- 170

QY 186 CRWGLAPREKLHCFLPQLAQISLIPDLFPRKALPD-----CEHA- 226

DB 171 --GVEDAQR-----RELGLPKASTPAPPRMAVRGSLAQVADALCFPL 212

QY 227 -----VGLRPOQPTG--SSTSYPPSPDKPRIFASLGTLO-GHYRGLFRTIAKAC 274

DB 213 AAEWGGRPFVFGALTWESATDADDEVASWIAADTPPIYFGSMPIGLSLADRVAMISAA 272

QY 275 EYVDAOLLAHCGLSATQAGLARGDQVVDFAQSAALSQAQLTTHGGMVTVLDAI 334

DB 273 AELGERALI--CSGPS--DATGIPQFDHVKVVRVSHAVFTCPRAVHHGGAGTTAAGL 328

QY 335 ASRTPLIALPLADPQGVASRIYHIGIGRASPRT--TSHALARQIRSLTITDYPQMT 392

DB 329 RAGIPTLLIWTSDQPIWAAQIKQLKVG-RGRFRSSATKESLIADRLTILA-PDYVTRAR 386

QY 393 KIOAAL-RLAGGTTPAAADIVQOAMR 416

DB 387 EIASRWTKPAASVTATADLLEDAAR 411

RESULT 12

UD88 RAT STANDARD; PRT; 530 AA.

AC Q62789;

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DE UDP-glucuronosyltransferase 2B8 precursor, microsomal (EC 2.4.1.17)

DE (UDPGR) (UGT2B-RH4).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

EN NCBI_TaxID=10116;

EN [1]

PP SEQUENCE FROM N.A.

RA STRAIN=HL2G-;

RA Cohen H., Trus M., Benvenisty N., Reshef L.;

RT "A novel member of the UDPGT family is abundantly expressed in H4IIEC3 hepatoma cells."

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC !- FUNCTION: UDPGT is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds

CC !- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

CC beta-D-glucuronoside.

CC !- SUBCELLULAR LOCATION: Microsomal.

CC !- SIMILARITY: Belongs to the UDP-glycosyltransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; U27518; AAA86833.1; -

DR

YP24 MYCTU STANDARD; PRT; 414 AA.

Q50583;

01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)

Hypothetical protein rv1524/MT1575/MD1551.

RV1524 OR MT1575 OR MTCY19C5.04C OR MB1551.

Mycobacterium tuberculosis, and

Mycobacterium bovis.

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

NCBI_TaxID=1773, 1765;

[1]

SEQUENCE FROM N.A.

SPECIES=M.tuberculosis; STRAIN=H37RV;

MDLINE=98295987; PubMed=9634230;

Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaija F., Badoock K., Basham D., Brown D., Chillingworth T., Connor R., Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagski K., Krogh A., Mclean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skellon S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."

Nature 393:537-544 (1998).

[2]

SEQUENCE FROM N.A.

SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;

MDLINE=22206494; PubMed=12218036;

Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;

"Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."

J. Bacteriol. 184:5479-5490 (2002).

[3]

SEQUENCE FROM N.A.

SPECIES=M.bovis; STRAIN=AF2122/97;

MDLINE=22709107; PubMed=12788972;

Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H., Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S., Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R., Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

"The complete genome sequence of Mycobacterium bovis."

Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).

!- SIMILARITY: STRONG, TO M.LEPRAE ML2348 AND M.TUBERCULOSIS

RV1526C.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; Z77826; CAB01393.1; -

EMBL; AB067024; AAK45842.1; -

EMBL; BX248339; CAD96218.1; -

PIR; D70723; D70723.

TIGR; MT1575; -

TubercuList; rv1524; -

InterPro; IPR004276; Glyco.trans.28.

Pfam; PF00303; Glyco.transf.28; 1.

Hypothetical protein; Complete proteome.

SEQUENCE 414 AA; 44361 MW; SDJ534C5173E2544 CRO64;

Query Match 7.1%; Score 158; DB 1; Length 414;

Best Local Similarity 24.3%; Pred. No. 7.1e-05;


```

MIM; 606434; -
MIM; 191740; -
GO; GO:0008152; p:metabolism; TAS.
InterPro; IPR002213; UDP_gluco_trans.
Pfam; PF00201; UDPGT; 1.
PROSITE; PS00375; UDPGT; 1.
Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
Multigene family; Microsome; Alternative splicing.
SIGNAL 1 25
CHAIN 26 530 UDP-GLUCURONOSYLTRANSFERASE 1-9.
TRANSMEM 488 504 POTENTIAL.
CARBOHYD 71 71 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 344 344 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 530 AA; 59940 MW; C417B9B86B403078 CRC64;

Query Match 6.7%; Score 150.5; DB 1; Length 530;
Best Local Similarity 19.4%; Pred. No. 0.00037;
Matches 91; Conservative 81; Mismatches 163; Indels 135; Gaps 20;

6 VIAPPPSHVRLAQLVARGHRTFFQCHDCKALVTGSDIGFO-----TVGLQTHP 60
29 LVVPMDSGFMTGRSVVEKILRGHEV-----VVVMPVSNQLGRSLNCTVKTY- 77
61 PGSLSHLH-----LAAHPLGPSMLRLINEMARTS-----DML-----CREL----- 97
78 --STSYTLEDLDRFKAFHQAQWKAQVRSIYSLMGSYNDIFDLFFSNCRSLPKDKKIVE 135
98 ---PAAFALQIBGVVDQMEPAGAVVAASGLPFVSV-----CALPLNRE 141
136 YLKESSFDA-----VFLPPDNCGLIVAKYFSLPSVVFARGILCHYLEGAQCPAPLSYV 190
142 PGLPLAVMPPEYGTSDAARERTTSEKIYDMLMRHRDVRVIAHACRMGLAPREKLHCFS 201
191 PRILL-----GFSDA-----MTFVRVNHIMLEHLLCHRRFFKNAIEASEILQ--T 237
202 PLAQ-----ISQLPELPPRKALPDCFHAVGLRQPOQTP----- 237
238 PVEYDLKSHTSIMLRTDFVLDYKPKVWPNMIP-IGSINCHQKGLPMBFEAYINASGE 296
238 -----GSSTSYFPPSPDKPRIFASLGTLOCHRYGLFRTIAKACEVDQALLAHCGGL 289
297 HGIWVFLSGSMVSEIPEKKAMAIADALG-----KIPQTVLWRYTG-- 336
290 SATQAGELARGDIOVDFADQSAALS--QAQLTIITGGMMTVLDAIASRTPLLALPLAF 347
337 --TRPSNLA--NNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMWMLFG 392
348 DQGVVASRIVVHGIGKASRP-TTSHALARQIRSLLTNTDYPQRMKIQ 396
393 DQMDNAKEMETKAGVTINLVENTSEDLNALKAVINDKSYKENIMRLSS 442

```

sarch completed: February 29, 2004, 14:45:06
 35 time : 10.264 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:51:24 ; Search time 28.6159 Seconds
(without alignments)
3180.293 Million cell updates/sec

file: US-09-941-947A-28

quence: 1 MSHFAVIAPPPFHHVRLQNL.....EQMRTCPVLSQDYATAL 431

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 211153259 residues

tal number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB	ID	Description
1	2231	100.0	431	10	US-09-941-947A-28	Sequence 28, Appl
2	2231	100.0	431	14	US-10-218-118-4	Sequence 4, Appli
3	251.5	11.3	397	14	US-10-156-761-14886	Sequence 14886, A
4	247	11.1	412	14	US-10-156-761-8483	Sequence 8483, Ap
5	177	7.9	477	9	US-09-740-029-4	Sequence 4, Appli
6	167	7.5	527	9	US-09-962-678-2	Sequence 2, Appli
7	167	7.5	527	9	US-09-981-353-166	Sequence 166, App
8	167	7.5	527	13	US-10-052-586-522	Sequence 522, App
9	167	7.5	527	14	US-10-174-590-522	Sequence 522, App
10	167	7.5	527	14	US-10-176-758-522	Sequence 522, App
11	167	7.5	527	14	US-10-175-737-522	Sequence 522, App
12	167	7.5	527	14	US-10-173-706-522	Sequence 522, App
13	167	7.5	527	14	US-10-175-738-522	Sequence 522, App
14	167	7.5	527	14	US-10-175-752-522	Sequence 522, App
15	167	7.5	527	14	US-10-176-482-522	Sequence 522, App

16	167	7.5	527	14	US-10-176-757-522	Sequence 522, App
17	167	7.5	527	14	US-10-176-913-522	Sequence 522, App
18	167	7.5	527	14	US-10-180-552-522	Sequence 522, App
19	167	7.5	527	14	US-10-180-557-522	Sequence 522, App
20	167	7.5	527	14	US-10-173-700-522	Sequence 522, App
21	167	7.5	527	14	US-10-174-572-522	Sequence 522, App
22	167	7.5	527	14	US-10-174-579-522	Sequence 522, App
23	167	7.5	527	14	US-10-174-582-522	Sequence 522, App
24	167	7.5	527	14	US-10-174-588-522	Sequence 522, App
25	167	7.5	527	14	US-10-175-739-522	Sequence 522, App
26	167	7.5	527	14	US-10-175-740-522	Sequence 522, App
27	167	7.5	527	14	US-10-175-743-522	Sequence 522, App
28	167	7.5	527	14	US-10-176-488-522	Sequence 522, App
29	167	7.5	527	14	US-10-176-492-522	Sequence 522, App
30	167	7.5	527	14	US-10-176-747-522	Sequence 522, App
31	167	7.5	527	14	US-10-176-750-522	Sequence 522, App
32	167	7.5	527	14	US-10-176-985-522	Sequence 522, App
33	167	7.5	527	14	US-10-176-987-522	Sequence 522, App
34	167	7.5	527	14	US-10-176-992-522	Sequence 522, App
35	167	7.5	527	14	US-10-176-993-522	Sequence 522, App
36	167	7.5	527	14	US-10-184-658-522	Sequence 522, App
37	167	7.5	527	14	US-10-176-991-522	Sequence 522, App
38	167	7.5	527	14	US-10-173-695-522	Sequence 522, App
39	167	7.5	527	14	US-10-173-697-522	Sequence 522, App
40	167	7.5	527	14	US-10-173-705-522	Sequence 522, App
41	167	7.5	527	14	US-10-174-576-522	Sequence 522, App
42	167	7.5	527	14	US-10-174-585-522	Sequence 522, App
43	167	7.5	527	14	US-10-174-586-522	Sequence 522, App
44	167	7.5	527	14	US-10-175-747-522	Sequence 522, App
45	167	7.5	527	14	US-10-176-481-522	Sequence 522, App

ALIGNMENTS

RESULT 1

US-09-941-947A-28
; Sequence 28, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-28

Query Match 100.0%; Score 2231; DB 10; Length 431;

Best Local Similarity 100.0%; Pred. No. 4.4e-218; Indels 0; Gaps 0;

Matches 431; Conservative 0; Mismatches 0;

QY 1 MSHFAVIAPPPFHHVRLQNLQELVARGHRTVFQOHDCKALVTGSDIGFTVGLQTHP 60

Db 1 MSHFAVIAPPPFHHVRLQNLQELVARGHRTVFQOHDCKALVTGSDIGFTVGLQTHP 60

QY 61 PGSLSHLLHLAAHPGLPGSMRLRLNEMARTSDMLCRELPAAHPALQIEGVIVDQEPAGAV 120

61 PGSLSHLLHAAHPLGSPMLRNLINENARTSDMLCRLPAALFALQIEGVIVDQMPAGAV 120
121 VAEASGLPVSFVACALPLNREPGPLVAVMPFEGTSDAAREYTTSEKIYDWMERHDRV 180
121 VAEASGLPVSFVACALPLNREPGPLVAVMPFEGTSDAAREYTTSEKIYDWMERHDRV 180
181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPKRALPCDFHAGVPLRQPGQTPGSS 240
181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPKRALPCDFHAGVPLRQPGQTPGSS 240
241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
301 GDIQVDFADQSAALSOAQLTTHGGMNTVLDIAASRTPLALPLAFDOPGVASRIVYHG 360
301 GDIQVDFADQSAALSOAQLTTHGGMNTVLDIAASRTPLALPLAFDOPGVASRIVYHG 360
361 IGRASRFTTSHALAQIRSLTNTDYPQRTMKIOAALRLAGGTPAAADIVEQAMRTQOP 420
361 IGRASRFTTSHALAQIRSLTNTDYPQRTMKIOAALRLAGGTPAAADIVEQAMRTQOP 420
421 VLSGQDYATAT 431
421 VLSGQDYATAT 431

RESULT 2
S-10-218-118-4
Sequence 4, Application US/10218118
Publication No. US20030148319A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
FILE REFERENCE: CL1876 US NA
CURRENT APPLICATION NUMBER: US/10/218,118
PRIOR FILING DATE: 2002-08-13
PRIOR FILING DATE: 2001-08-15
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 431
TYPE: PRT
ORGANISM: Pantoea stewartii
S-10-218-118-4

Query Match 100.0%; Score 2231; DB 14; Length 431;
Best Local Similarity 100.0%; Pred. No. 4,4e-218;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSHFAVIAPPPFSHVRLALQALVARGHRTFFQOHDCKALVTGSDIGFTVGLQTHP 60
1 MSHFAVIAPPPFSHVRLALQALVARGHRTFFQOHDCKALVTGSDIGFTVGLQTHP 60
61 PGSLSHLLHAAHPLGSPMLRNLINENARTSDMLCRLPAALFALQIEGVIVDQMPAGAV 120
61 PGSLSHLLHAAHPLGSPMLRNLINENARTSDMLCRLPAALFALQIEGVIVDQMPAGAV 120
121 VAEASGLPVSFVACALPLNREPGPLVAVMPFEGTSDAAREYTTSEKIYDWMERHDRV 180
121 VAEASGLPVSFVACALPLNREPGPLVAVMPFEGTSDAAREYTTSEKIYDWMERHDRV 180
181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPKRALPCDFHAGVPLRQPGQTPGSS 240
181 IAHACRMGLAPREKLHCHFCSPLAQISQILPELDPKRALPCDFHAGVPLRQPGQTPGSS 240
241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300

241 TSYPSPDKPRIFASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
301 GDIQVDFADQSAALSOAQLTTHGGMNTVLDIAASRTPLALPLAFDOPGVASRIVYHG 360
301 GDIQVDFADQSAALSOAQLTTHGGMNTVLDIAASRTPLALPLAFDOPGVASRIVYHG 360
361 IGRASRFTTSHALAQIRSLTNTDYPQRTMKIOAALRLAGGTPAAADIVEQAMRTQOP 420
361 IGRASRFTTSHALAQIRSLTNTDYPQRTMKIOAALRLAGGTPAAADIVEQAMRTQOP 420
421 VLSGQDYATAT 431
421 VLSGQDYATAT 431

RESULT 3
US-10-156-761-14886
Sequence 14886, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHARA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14886
LENGTH: 397
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14886

Query Match 11.3%; Score 251.5; DB 14; Length 397;
Best Local Similarity 26.6%; Pred. No. 1.2e-16;
Matches 115; Conservative 61; Mismatches 186; Indels 71; Gaps 18;

1 MSHFAVIAPPPFSHVRLALQALVARGHRTFFQOHDCKALVTGSDIGFTVGLQTHP 60
1 MGRFLFVVPPLVGHVNPVAVCTAAALARGHDIAGHGPVLVGLAGADAWFPCAL---P 57
61 PGSLSHLLHAAHPLGSPMLRNLINENARTSDMLCRLPAALFALQIEGVIVDQMPAG 118
58 EDGLSRPAGLK----GPAAFQFLWESFLVPLADAMAPGVRAAIEAYDPDVVYCDQAVAG 113
119 AVVAEASGLPVSFVACALPLNREPGPLVAVMPFEGTSDAAREYTTSEKIYDWMERH 178
114 ALVAESGLPVSFVACALPLNREPGPLVAVMPFEGTSDAAREYTTSEKIYDWMERH 149
179 RVIAHACRM---GLA--PREKLHCHFCSPLAQISQILPELDPKRALPCDFHAGV--LR 231
150 GLLGELRRRITGAGAADPR-----FSPHGVLAIVTTRALLGPVE--LPDRVWLVGPSVAA 202
232 OPQGTGSGSTSFPSDPKPRIPASLGTLOGHRYGLFRTIAKACEVDQAQLLAHCGGLS- 290
203 RPAGPDOPFWMELEASALPTVLVSLGTANNDAGAFNLNA-----AEAL-----GGIAD 251
291 ATQAGELARGGDIQ-----VVPDQSAALSOAQLTTHGGMNTVLDIAASRTPLLA 342
252 RVRAVLVDPGVVEHPVDPDTLVRRVYVQLALLERLDAVVCAGHNTVCEALWHGVPLV 311
343 LPALFOPGVASRIVYHGIGKASRTTSHA--LARGISLITNTDYPQRTMK-1QAALR 399
312 AP-RRDQDPITVAQVVDAGAGVRL-RFRGADAARIAGAAVEAVLDPAQGHKRAAAGVBSFR 370

Patent No. US20020082194A1
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
DRUG-METABOLIZING PROTEINS,
AND USES THEREOF
FILE REFERENCE: CLO00928
CURRENT APPLICATION NUMBER: US/09/740,029
CURRENT FILING DATE: 2000-12-20
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 477
TYPE: PRT
ORGANISM: Oryctolagus cuniculus
US-09-740-029-4

Query Match 7.9%; Score 177; DB 9; Length 477;
Best Local Similarity 21.5%; Pred. No. 6.1e-09;
Matches 96; Conservative 75; Mismatches 177; Indels 98; Gaps 19;

QY 13 SHVRALQNLAEVLVARGHRTFFQCHDKALVTGSDIGFQTVGLQTHPP-----GSLSHL 67
DB 1 SHWINLKVLEELQURGHRTVLPSP-SLLDHTKIPFNVEVLQVTKETLMEEELNTV 59
QY 68 LHLAAHPGLPSM-----LRLINEMWARTSDMLCRELPAAFAHALQIEG 108
DB 60 LYMSSFEL-PTLSWTKVLGMVEMGKQPSKMLRRVCDSAITNKELDLXAA-----KFDI 114
QY 109 VIVDQMEPAGAVVAEASGLPPV-----SVACALPLNRPGPLAVMPFEY--GT 155
DB 115 CLADPLAPCGELVAELNIPVYSFRPSIGNIIRSCA-----GLP--TPSSYVPGS 164
QY 156 SDABERVTSEKIYDML-RRHDEVIAH-----ACR-MCLAPREK 195
DB 165 TSGUTDNMSFVORLKNWLLYLMNDMFSHMLSWDEYYSKVLGRRTTICEIMCKAEWL 224
QY 196 LHHCFSPLAQISQLPELDPKALPDCFHAVGEL--RQPGTSGSTSTVPPSP--DKPR 251
DB 225 IRSYV-----DEEFPRLPN-FEYVGLHCKPAKPLPEELEEFVQSSGNDGW 272
QY 252 IFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARGDIOVVDFADQ 311
DB 273 VFTLGSMTQNLTEERSNLIALAQIPQKVLWRVTGKKPAT-----LGNTRLFEWIFQ 326
QY 312 SAALS--QAQLTIHGGMTVLDIASRTPLLLALPLAPDQPGVASRIVYHIGIKRAS-RF 368
DB 327 NDLGHKPTRAFITHGGTNGLYEALYHGVPMVGIPLFGDQPDNIARVKAKGAADVDLRI 386
QY 369 TTSALARQIRSLTNTDYPORMTKI 394
DB 387 MTTSSLKALKADVINNPYSKENAMKL 412

RESULT 6
US-09-962-678-2
Sequence 2, Application US/09962678
Patent No. US20020155499A1
GENERAL INFORMATION:
APPLICANT: Leiby, Kevin R.
TITLE OF INVENTION: 32624, A NOVEL HUMAN UDP-GLUCURONOSYL
TRANSFERASE AND USES THEREOF
FILE REFERENCE: 10448-094001
CURRENT APPLICATION NUMBER: US/09/962,678
CURRENT FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/235,044
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 527
TYPE: PRT

400 LAGCTPAAADIVE 412
|||:||||:
371 AAGGSAADRL 383

SULT 4
-10-156-761-8483
Sequence 8483, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8483
LENGTH: 412
TYPE: PRT
ORGANISM: Streptomyces avermitilis
-10-156-761-8483

Query Match 11.1%; Score 247; DB 14; Length 412;
Best Local Similarity 21.3%; Pred. No. 3.7e-16;
Matches 102; Conservative 76; Mismatches 162; Indels 138; Gaps 15;

3 HEVATPFFSHVRALQNLAEVLVARGHRTFFQCHDKALVTGSDI-----49
4 HFLFMSAPFGHVPFSLVAELVHRGHVTF-----VTGAEXADAVRSVGADFLR 54
50 ---GFQTVGLQTHPPGSLHLLAAHPLGSPMLRLINEMA--RTSDMLCRELPAAFAHAL 104
55 YESAFEGVDYRLMTAEAPNAIPMTLYDEGMSMLRSVEEFGKDVPLVAYDIATSLN-- 112
105 QLEGVIVDQMEPAGAVVAEASGLFPVSACALPLNRPGPLAVMPFEYGTSDAARERT 164
113 -----VGRVLAASWSRPAMTV-----IPLF-----ASNGRFS 139
165 TSEKIYDMLRRHDRVIAHACRMGLAPREKLHHCFS---PLAQISQLPEL----- 213
140 TMQSVLD-----PDSAQVSAPPFR-----FSQEMELFGLGALVPRLAELLVSRG 183
214 -----DPPRKALPDCFH-----AVGPLRQPGTSGSTSTVPPSPDK 249
184 ITEPVDDFLSGPEDFNLVCLPRAFOYAGDTDFERPAFVGPCLGKRGGLGEWTP--PGSGH 241
250 PRIFASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGLSATQAGELARGD----- 302
242 PVVLISLGTVFNQRSLSPFRIFVRAETDVPVHVVIS-----LKGVDVPLRPL 289
303 ---IQVDFADQSAALSQAQLTIHGGMTVLDIASRTPLLLALPLAPDQPGVASRIVYH 359
290 PENVEVHRVPHAVLEHARALVTHGGTGVMEALHAGCPVLVMPLSRDAQVTRIAEL 349
360 GIGKAS-RFTTSHALARQIRSLTNTDYPORMTKIOAALRLAGGTPAAADIVEQAMR 416
350 GLGRMVQPEVTTATLRRHVLDDIISDDAITROVQMRQTVEAGGALRAADETERFLR 407

350 LAGCTPAAADIVE 412
|||:||||:
371 AAGGSAADRL 383

SULT 5
3-09-740-029-4
Sequence 4, Application US/09740029

ORGANISM: Homo sapiens

IS-09-962-678-2

Query Match 7.5%; Score 167; DB 9; Length 527;

Best Local Similarity 20.8%; Pred. No. 7.4e-08;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

6 VIAPPPFHVRLQNLQALQELVARGHRTVFPQDCKALV---TGSDIGFQTVGLQTHPPG 62

27 LWVPCDMSHMLNKKVLEELIVRGHEVTVL-TSKPSLIDYRKPSALKFEVVMQDRT 85

63 SLSHLLHAAHPL-----GFSMLRLIN---EMARTSDMLC-----RELPAAPHALQIEG 108

86 ENEIFVDLALNVLPLGLSTWQSVIKLNDFFVEIRGTLLKMCESFYINQTLMKKLQETNYDV 145

109 VIVQMPAGAVARASGLPFV---SVACALPLNRPB---LPLAVMPPEY-GTSDAARE 161

146 MLIDPVIPCGDLMAELLAVFVLTLRISVGNMERSCGKLPAPLSYVVPVMTGLTD---- 201

162 RYTTSEKI-----YD-----WDFEFPQYQPN-FEFVGLHCKPAKALP 287

202 RMTFLERVKNMSLVLFHFHFIQDYHYHFEWFYSKALGRPTTLCTETVGKAEIWLIRTY-- 259

180 VIAHACRMGLAPREKHLHCFSPLAQISQLIPDLDFPRKALPDCFHAVGPEL--RQPGQTP 237

260 -----WDFEFPQYQPN-FEFVGLHCKPAKALP 287

238 GSSTSYFSPDKPRIPA-SLGTLL-QCHRYGLFRTIAKACEVDQAQLLHACGGLSATQAG 295

288 KEMENFVQSSGEDIIVFSLGSLFQNVTEEKANIIASALAQIPQKVLRYKGGKPS--- 344

296 ELARGDGIQVDFADQSAALS---QAOLTIHGGNTVLDIAISRTPLLAFLAFDQ----- 349

345 ---LGNATRLYDWTIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMGVPIFGDQDNI 401

350 -----PGVASRIVVHGIGKASRPTTSHALAROIRSLTNTDYPQMTKIQAAALRLA 401

402 AHMKAKGAAVEINP-----KTMISEDLLRALRTVITDSSYKEN-----AMRLS 444

RESULT 7

US-09-981-353-166

Sequence 166, Application US/09981353

Patent No. US20020160382A1

GENERAL INFORMATION:

APPLICANT: Lasek, Amy W.

APPLICANT: Jones, David A.

TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

FILE REFERENCE: PA-0038 US

CURRENT APPLICATION NUMBER: US/09/981,353

CURRENT FILING DATE: 2001-10-11

NUMBER OF SEQ ID NOS: 194

SOFTWARE: PERL Program

SEQ ID NO 166

LENGTH: 527

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. US20020160382A1 2434655CD1

US-09-981-353-166

Query Match

Best Local Similarity 20.8%; Score 167; DB 9; Length 527;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

6 VIAPPPFHVRLQNLQALQELVARGHRTVFPQDCKALV---TGSDIGFQTVGLQTHPPG 62

27 LWVPCDMSHMLNKKVLEELIVRGHEVTVL-TSKPSLIDYRKPSALKFEVVMQDRT 85

63 SLSHLLHAAHPL-----GFSMLRLIN---EMARTSDMLC-----RELPAAPHALQIEG 108

86 ENEIFVDLALNVLPLGLSTWQSVIKLNDFFVEIRGTLLKMCESFYINQTLMKKLQETNYDV 145

PRIOR	FILING DATE: 1998-05-07	
PRIOR	APPLICATION NUMBER: 60/085573	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085579	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085580	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085582	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/085700	
PRIOR	FILING DATE: 1998-05-15	
PRIOR	APPLICATION NUMBER: 60/086023	
PRIOR	FILING DATE: 1998-05-18	
PRIOR	APPLICATION NUMBER: 60/086392	
PRIOR	FILING DATE: 1998-05-22	
PRIOR	APPLICATION NUMBER: 60/086486	
PRIOR	FILING DATE: 1998-05-22	
PRIOR	APPLICATION NUMBER: 60/087098	
PRIOR	FILING DATE: 1998-05-28	
PRIOR	APPLICATION NUMBER: 60/087208	
PRIOR	FILING DATE: 1998-05-28	
PRIOR	APPLICATION NUMBER: 60/087609	
PRIOR	FILING DATE: 1998-06-02	
PRIOR	APPLICATION NUMBER: 60/087759	
PRIOR	FILING DATE: 1998-06-02	
PRIOR	APPLICATION NUMBER: 60/087827	
PRIOR	FILING DATE: 1998-06-03	
PRIOR	APPLICATION NUMBER: 60/088025	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088028	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088029	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088033	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088167	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088202	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088212	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088217	
PRIOR	FILING DATE: 1998-06-05	
PRIOR	APPLICATION NUMBER: 60/088326	
PRIOR	FILING DATE: 1998-06-04	
PRIOR	APPLICATION NUMBER: 60/088555	
PRIOR	FILING DATE: 1998-06-09	
PRIOR	APPLICATION NUMBER: 60/088722	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088738	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088740	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088811	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088824	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088825	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088826	
PRIOR	FILING DATE: 1998-06-10	
PRIOR	APPLICATION NUMBER: 60/088861	
PRIOR	FILING DATE: 1998-06-11	
PRIOR	APPLICATION NUMBER: 60/088863	
PRIOR	FILING DATE: 1998-06-11	
PRIOR	APPLICATION NUMBER: 60/088876	
PRIOR	FILING DATE: 1998-06-11	
PRIOR	APPLICATION NUMBER: 60/089090	
PRIOR	FILING DATE: 1998-06-12	
PRIOR	APPLICATION NUMBER: 60/089105	
PRIOR	FILING DATE: 1998-06-12	
PRIOR	APPLICATION NUMBER: 60/089512	
PRIOR	FILING DATE: 1998-06-16	

PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089908

Query Match 7.5%; Score 167; DB 13; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;
Y 6 VIAPPPFHVRLONLAQELVARGHVTFQQHCKALV---TGSDIGFQTVGLQTHPPG 62
D 27 LVNFCDSHNLNKKVLEELVIRGHEVTVL---THSKPSLDYRKPSALKFEVHMPQDRTE 85
Y 63 SLSHLLHAAHPL-----GPSMLRLN---EMARTSDMLC-----RELPAAFHALQIEG 108
D 86 ENEIFVDLALNVLPGSLSTQSVIKLNDFFVEIRGTLKMCESFIYNQTLMKKLOETNYDV 145
Y 109 VVDQMPAGAVVAEASGLPFV---SVACALPLNREP---LPLAVMPPEY-GTSDAARE 161
D 146 MLIDPVIPCGDLMAELLAVPFVLTLRISVGNMERSCGKLPAPLSYVPVMTGLTD---- 201
Y 162 RYTTSSEKI-----YD-----WLMRHRDR 179
D 202 RMTFLERKVMNSLVLFHFWIQDYDHFWEFYSKALGRPTTLCTVGAETWIRTY--- 259
Y 180 VIAHACRMGLAPREKLHCHFCSPAQISQLIPBLDPRKALPDCFEAVGLP---RPOQGT 237
D 260 -----WDFEFPQYQPN---PEFVGLHCKPAKALP 287
Y 238 GSSTSYFSPDPKPRIFA-SLGTIL-QGHRYGLFRTIAKACEVDQAQLLHACGSLSATQAG 295
D 288 KEMENFVQSSGEGIVVFSIGSLFQNVTEKANIASALAQIPQKVLWRYGKKPST--- 344
Y 296 ELARGGDIQVDFADQSAALS---QAQUTITHGWNVTDLAASRTPELLALPLAFDQ--- 349
D 345 ---LGANTRYLDYDIPQNDLLGHPKTRAFITHGGMNGIYEALYHGVPMVGVPIFGDQDNI 401
Y 350 -----PGVASRIVHVGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
D 402 AHMKAKGAARVEINF-----KMTSDDLRLALRTVITDSSYKEN-----AMRLS 444

RESULT 9
US-10-174-590-522
Sequence 522, Application US/10174590
Publication No. US20030008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42
CURRENT APPLICATION NUMBER: US/10/174,590
CURRENT FILING DATE: 2002-06-18
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT

ORGANISM: Homo Sapien
US-10-174-590-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;
QY 6 VIAPPPFHVRLONLAQELVARGHVTFQQHCKALV---TGSDIGFQTVGLQTHPPG 62
DB 27 LVNFCDSHNLNKKVLEELVIRGHEVTVL---THSKPSLDYRKPSALKFEVHMPQDRTE 85
QY 63 SLSHLLHAAHPL-----GPSMLRLN---EMARTSDMLC-----RELPAAFHALQIEG 108
DB 86 ENEIFVDLALNVLPGSLSTQSVIKLNDFFVEIRGTLKMCESFIYNQTLMKKLOETNYDV 145
QY 109 VVDQMPAGAVVAEASGLPFV---SVACALPLNREP---LPLAVMPPEY-GTSDAARE 161
DB 146 MLIDPVIPCGDLMAELLAVPFVLTLRISVGNMERSCGKLPAPLSYVPVMTGLTD---- 201
QY 162 RYTTSSEKI-----YD-----WLMRHRDR 179
DB 202 RMTFLERKVMNSLVLFHFWIQDYDHFWEFYSKALGRPTTLCTVGAETWIRTY--- 259
QY 180 VIAHACRMGLAPREKLHCHFCSPAQISQLIPBLDPRKALPDCFEAVGLP---RPOQGT 237
DB 260 -----WDFEFPQYQPN---PEFVGLHCKPAKALP 287
QY 238 GSSTSYFSPDPKPRIFA-SLGTIL-QGHRYGLFRTIAKACEVDQAQLLHACGSLSATQAG 295
DB 288 KEMENFVQSSGEGIVVFSIGSLFQNVTEKANIASALAQIPQKVLWRYGKKPST--- 344
QY 296 ELARGGDIQVDFADQSAALS---QAQUTITHGWNVTDLAASRTPELLALPLAFDQ--- 349
DB 345 ---LGANTRYLDYDIPQNDLLGHPKTRAFITHGGMNGIYEALYHGVPMVGVPIFGDQDNI 401
QY 350 -----PGVASRIVHVGIGKASRFTTSHALAROIRSLTNTDYPQRMTKIQALRLA 401
DB 402 AHMKAKGAARVEINF-----KMTSDDLRLALRTVITDSSYKEN-----AMRLS 444

RESULT 10
US-10-176-758-522
Sequence 522, Application US/10176758
Publication No. US20030008353A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C104
CURRENT APPLICATION NUMBER: US/10/176,758
CURRENT FILING DATE: 2002-06-21
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-758-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

```

Y 6 VIAPFFSHVRLQNLQAEVLVARGHRVTFQHQDCKALV---TGSDIGFTVGLQTHPPG 62
b 27 LVWPCDMSHNLVAVKLELIVRGHEVTVL--THSKPSLIDYKPSALKFVYVHPQDRTE 85
Y 63 SLSHLLHLAAHPL-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108
b 86 ENEIFVDLALNVLPGSLTWQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145
Y 109 VIVDQMEPAGVVAEASGLPFV---SVACALPLNREP---LPLAMPPEY-GTSDAARE 161
b 146 MLIDPVIPOCDLMAELLAVFVLTIRISVGGNMERSCGKLPAPLSVYVPMVMTGLTD----- 201
Y 162 RYTTSEKI-----YD-----WLMRHRDR 179
b 202 RMTFLERVNSMLSVLFHFHWIQQDYHFWEEFYSKALGRPTTLCTETVGKAEIWLIRTY-- 259
Y 180 VIAHACRMGLAPREKLHCHFCPSLAQISOLIPELDPKPKALPDCEHVGPL--RPOQGTTP 237
b 260 -----WDFBFPQPYQPN--FEFVGGHLCKPAKALP 287
Y 238 GSSTSYFPPSPDKPRIPA-SLGTIL-QGHRVGLFRTITAKACEEVDQAOLLAHCGGLSATQAG 295
b 288 KENENFVQSGGEGIVVFSLSLQFNVTTEKANIITASALAQIPQKVLWRYKCKKPEST--- 344
Y 296 ELARGDIOVDFADQSAALS--QAOLTIITHGGMNTVLDIASRTPPLALPLAFDQ----- 349
b 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIVHGVPMVGVPIFGDQLDNI 401
Y 350 -----PGVASRIVYHIGIKRASRFTTSHALAQIRSLTNTDYPQRMWKIOAALRLA 401
b 402 AHMKAGRAVEINF-----KMTSEDLRLALRTVITDSSYKEN-----AMRLS 444

```

ESULT 11

S-10-175-737-522

Sequence 522, Application US/10175737

Publication No. US20030013153A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430RIC50

CURRENT APPLICATION NUMBER: US/10/175,737

CURRENT FILING DATE: 2002-06-19

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 522

LENGTH: 527

TYPE: PRT

ORGANISM: Homo Sapien

S-10-175-737-522

Query Match 7.5%; Score 167; DB 14; Length 527;

Best Local Similarity 20.8%; Pred. No. 7.4e-08;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

Y 6 VIAPFFSHVRLQNLQAEVLVARGHRVTFQHQDCKALV---TGSDIGFTVGLQTHPPG 62

b 27 LVWPCDMSHNLVAVKLELIVRGHEVTVL--THSKPSLIDYKPSALKFVYVHPQDRTE 85

Y 63 SLSHLLHLAAHPL-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108

b 86 ENEIFVDLALNVLPGSLTWQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145

Y 109 VIVDQMEPAGVVAEASGLPFV---SVACALPLNREP---LPLAMPPEY-GTSDAARE 161

b 146 MLIDPVIPOCDLMAELLAVFVLTIRISVGGNMERSCGKLPAPLSVYVPMVMTGLTD----- 201

Y 162 RYTTSEKI-----YD-----WLMRHRDR 179

b 202 RMTFLERVNSMLSVLFHFHWIQQDYHFWEEFYSKALGRPTTLCTETVGKAEIWLIRTY-- 259

Y 180 VIAHACRMGLAPREKLHCHFCPSLAQISOLIPELDPKPKALPDCEHVGPL--RPOQGTTP 237

b 260 -----WDFBFPQPYQPN--FEFVGGHLCKPAKALP 287

Y 238 GSSTSYFPPSPDKPRIPA-SLGTIL-QGHRVGLFRTITAKACEEVDQAOLLAHCGGLSATQAG 295

b 288 KENENFVQSGGEGIVVFSLSLQFNVTTEKANIITASALAQIPQKVLWRYKCKKPEST--- 344

Y 296 ELARGDIOVDFADQSAALS--QAOLTIITHGGMNTVLDIASRTPPLALPLAFDQ----- 349

b 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIVHGVPMVGVPIFGDQLDNI 401

Y 350 -----PGVASRIVYHIGIKRASRFTTSHALAQIRSLTNTDYPQRMWKIOAALRLA 401

b 402 AHMKAGRAVEINF-----KMTSEDLRLALRTVITDSSYKEN-----AMRLS 444

```

Y 109 VIVDQMEPAGVVAEASGLPFV---SVACALPLNREP---LPLAMPPEY-GTSDAARE 161
b 146 MLIDPVIPOCDLMAELLAVFVLTIRISVGGNMERSCGKLPAPLSVYVPMVMTGLTD----- 201
Y 162 RYTTSEKI-----YD-----WLMRHRDR 179
b 202 RMTFLERVNSMLSVLFHFHWIQQDYHFWEEFYSKALGRPTTLCTETVGKAEIWLIRTY-- 259
Y 180 VIAHACRMGLAPREKLHCHFCPSLAQISOLIPELDPKPKALPDCEHVGPL--RPOQGTTP 237
b 260 -----WDFBFPQPYQPN--FEFVGGHLCKPAKALP 287
Y 238 GSSTSYFPPSPDKPRIPA-SLGTIL-QGHRVGLFRTITAKACEEVDQAOLLAHCGGLSATQAG 295
b 288 KENENFVQSGGEGIVVFSLSLQFNVTTEKANIITASALAQIPQKVLWRYKCKKPEST--- 344
Y 296 ELARGDIOVDFADQSAALS--QAOLTIITHGGMNTVLDIASRTPPLALPLAFDQ----- 349
b 345 ---LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIVHGVPMVGVPIFGDQLDNI 401
Y 350 -----PGVASRIVYHIGIKRASRFTTSHALAQIRSLTNTDYPQRMWKIOAALRLA 401
b 402 AHMKAGRAVEINF-----KMTSEDLRLALRTVITDSSYKEN-----AMRLS 444

```

RESULT 12

US-10-173-706-522

Sequence 522, Application US/10173706

Publication No. US20030022293A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCES: P3430RIC7

CURRENT APPLICATION NUMBER: US/10/173,706

CURRENT FILING DATE: 2002-06-17

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 522

LENGTH: 527

TYPE: PRT

ORGANISM: Homo Sapien

US-10-173-706-522

Query Match 7.5%; Score 167; DB 14; Length 527;

Best Local Similarity 20.8%; Pred. No. 7.4e-08;

Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

Y 6 VIAPFFSHVRLQNLQAEVLVARGHRVTFQHQDCKALV---TGSDIGFTVGLQTHPPG 62

b 27 LVWPCDMSHNLVAVKLELIVRGHEVTVL--THSKPSLIDYKPSALKFVYVHPQDRTE 85

Y 63 SLSHLLHLAAHPL-----GPSMLRLIN---EMARTSDMLC-----RELPAAFHALQIEG 108

b 86 ENEIFVDLALNVLPGSLTWQSVIKLNDFFVEIRGTLKWCESFIYNQTLMKKLOETNYDV 145

Y 109 VIVDQMEPAGVVAEASGLPFV---SVACALPLNREP---LPLAMPPEY-GTSDAARE 161

b 146 MLIDPVIPOCDLMAELLAVFVLTIRISVGGNMERSCGKLPAPLSVYVPMVMTGLTD----- 201

Y 162 RYTTSEKI-----YD-----WLMRHRDR 179

```

202 RMTFLERVKNSMLSVLFHFWIQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
180 VIAHACRMGLAPREKLHCFSPLAQISQLIPELDFPRKALPDCFHAVGEL--RQPQGT 237
260 -----WDFEPQPYQPN-FEFVGLHCKPAKALP 287
238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEEKANIISALAQIPQKVLWRYGKKPST--- 344
296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
345 ---LGANTRLYDWIPQNDLLGHKPKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
350 -----PCVASRIYVHGIKGRASRTTSHALAROIRSLTNTDYPQRTKIQAALRLA 401
402 AHMKAGAAVEINF-----KTMTSDELRLRTRVITDSSYKEN-----AMRLS 444

RESULT 13
US-10-175-738-522
Sequence 522, Application US/10175738
Publication No. US2003002294A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C45
CURRENT APPLICATION NUMBER: US/10/175,738
CURRENT FILING DATE: 2002-06-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-738-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

202 RMTFLERVKNSMLSVLFHFWIQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
180 VIAHACRMGLAPREKLHCFSPLAQISQLIPELDFPRKALPDCFHAVGEL--RQPQGT 237
260 -----WDFEPQPYQPN-FEFVGLHCKPAKALP 287
238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEEKANIISALAQIPQKVLWRYGKKPST--- 344
296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
345 ---LGANTRLYDWIPQNDLLGHKPKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
350 -----PCVASRIYVHGIKGRASRTTSHALAROIRSLTNTDYPQRTKIQAALRLA 401
402 AHMKAGAAVEINF-----KTMTSDELRLRTRVITDSSYKEN-----AMRLS 444

RESULT 14
US-10-175-752-522
Sequence 522, Application US/10175752
Publication No. US2003002295A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60
CURRENT APPLICATION NUMBER: US/10/175,752
CURRENT FILING DATE: 2002-06-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-752-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

202 RMTFLERVKNSMLSVLFHFWIQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
180 VIAHACRMGLAPREKLHCFSPLAQISQLIPELDFPRKALPDCFHAVGEL--RQPQGT 237
260 -----WDFEPQPYQPN-FEFVGLHCKPAKALP 287
238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEEKANIISALAQIPQKVLWRYGKKPST--- 344
296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
345 ---LGANTRLYDWIPQNDLLGHKPKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
350 -----PCVASRIYVHGIKGRASRTTSHALAROIRSLTNTDYPQRTKIQAALRLA 401
402 AHMKAGAAVEINF-----KTMTSDELRLRTRVITDSSYKEN-----AMRLS 444

```

```

288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEEKANIISALAQIPQKVLWRYGKKPST--- 344
296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
345 ---LGANTRLYDWIPQNDLLGHKPKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
350 -----PCVASRIYVHGIKGRASRTTSHALAROIRSLTNTDYPQRTKIQAALRLA 401
402 AHMKAGAAVEINF-----KTMTSDELRLRTRVITDSSYKEN-----AMRLS 444

RESULT 14
US-10-175-752-522
Sequence 522, Application US/10175752
Publication No. US2003002295A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60
CURRENT APPLICATION NUMBER: US/10/175,752
CURRENT FILING DATE: 2002-06-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-752-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;

202 RMTFLERVKNSMLSVLFHFWIQDYHFWEEFYSKALGRPTTLCETVGKAEIWLRTY--- 259
180 VIAHACRMGLAPREKLHCFSPLAQISQLIPELDFPRKALPDCFHAVGEL--RQPQGT 237
260 -----WDFEPQPYQPN-FEFVGLHCKPAKALP 287
238 GSSTSYFSPDKPRIFA-SLGTLL-QGHRYGLFRTIAKACEEVDQAQLLAHCCGLSATQAG 295
288 KEMENFVQSSGEDIIVFVLSGSLFQNTVEEKANIISALAQIPQKVLWRYGKKPST--- 344
296 ELARGDGIQVDFADQSAALS--QAQLTIHGGMNTVLDIASRTPLALPLAFDQ---- 349
345 ---LGANTRLYDWIPQNDLLGHKPKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQDNI 401
350 -----PCVASRIYVHGIKGRASRTTSHALAROIRSLTNTDYPQRTKIQAALRLA 401
402 AHMKAGAAVEINF-----KTMTSDELRLRTRVITDSSYKEN-----AMRLS 444

```

350 -----PGVASRIYVHGIGKRSRFTTSHALAQIRSLTNTDYPQRTKIQALRLA 401
402 AHMKAKGAARVEINF-----KWTSEDLLRALRTVTITDSSYKEN-----AMRLS 444

RESULT 15

-10-176-482-522
Sequence 522, Application US/10176482
Publication No. US20030022296A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C70
CURRENT APPLICATION NUMBER: US/10/176,482
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
-10-176-482-522

Query Match 7.5%; Score 167; DB 14; Length 527;
Best Local Similarity 20.8%; Pred. No. 7.4e-08;
Matches 99; Conservative 79; Mismatches 159; Indels 140; Gaps 21;
6 VIAPFFSHVRALQNLQELVARGHRVTFQOHCKALV---TGSIDIGFQTVGLQTHPPG 62
27 LVWPCDMSHMLNVKVLSELIVRGHEVTVL-THSKSLIDYRKPSALKFEVWHMPQDRTE 85
63 SLSELHLHAALPL-----GPSMLRLIN---EMARTSDMLC-----RELPAAPHALQIEG 108
86 ENEIFVDLAINVLPGLSWQSVIKLNDFFVEIRGLTKMWCESFIYNQTLMKKLOETNYDV 145
109 VIVDQMEPAGAVVAEASGLPFV---SVACALPLNREPG---LPLAVMPFEY-GTSDAARE 161
146 MLIDSVIFCGDLMAELAVPFVLTIRISVGNMERSCKLPAPLSYVPEVMTGLTD---- 201
162 RYTTSEKI-----YD-----WLMRRHDR 179
202 RMTFLERVKNMLSVLHFHFIQDDYHFWEFYSKALGRPTTLCETVGKAEIWLIRTY-- 259
180 VIAHACRMGLAPREKLHRCFSLAQISQLIPELDFPRKALPDCFHAVGPL--RQPGGTP 237
260 -----WDFEFPQYQPN-FEFVGGFLCKKPAKALP 287
238 GSSTSYFSPDKPRIFA-SIGTL-QCHRYGLFRIAKACEVDQAQLLHACGGLSATQAG 295
288 KEMENFVQSSGSDGIVWFSLGSLFQNTTEKANTIASALAQIPOKVLRYRYKKPST--- 344
296 ELARGDITQWVDFADQSAALS--QAOLTIHGGNTVLDALIASRTPLIALPLAFDQ---- 349
345 ---LQANTRLYDWIFQNDLLGHPTKAPITHGGNGNIVEALYHGVPMVGVPIFGDQDNI 401
350 -----PGVASRIYVHGIGKRSRFTTSHALAQIRSLTNTDYPQRTKIQALRLA 401
402 AHMKAKGAARVEINF-----KWTSEDLLRALRTVTITDSSYKEN-----AMRLS 444

Search completed: February 29, 2004, 15:27:58
Job time : 32.6159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

in on: February 29, 2004, 14:34:14 ; Search time 11.9967 Seconds
(without alignments)
3455.835 Million cell updates/sec

File: US-09-941-947a-28
Project score: 2231
Sequence: 1 MSHFAVIAPPPFHVRLQN.....EQAMRTQPVLSGGDYATAL 431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1794	80.4	431	2 S52583	ctxX protein - Erw
2	1790	80.2	431	2 B37802	ctxX protein - Erw
3	1018.5	45.7	413	2 S52980	zeaxanthin glucosyl
4	558.5	25.0	419	1 S74500	zeaxanthin glucosyl
5	262	11.7	407	2 T37104	probable glycosyl
6	228.5	10.2	430	1 S33184	glycosyltransferase
7	217	9.7	424	2 T51110	glycosyltransferase
8	204	9.1	392	1 C69851	macrolide glycosyl
9	198.5	8.9	418	2 F75887	probable glycosyl
10	195.5	8.8	287	2 H69306	macrolide glycosyl
11	181.5	8.1	418	2 JS0636	glycosyl transferase
12	180	8.1	54	2 A46292	zeaxanthin glucosyl
13	175.5	7.9	541	2 JCS423	2-hydroxyacylshpin
14	175	7.8	407	2 C92423	probable transferase
15	172	7.7	388	2 B70878	probable glycosyl
16	168.5	7.6	411	2 T03890	hypothetical prote
17	167	7.5	455	1 XUBHFG	flavonol 3-O-gluc
18	164.5	7.4	541	2 A48801	2-hydroxyacylshpin
19	164	7.4	392	2 S72936	UDP-glucuronosyltr
20	162	7.3	535	2 I57961	glucuronosyltransf
21	160.5	7.2	449	2 G70670	hypothetical prote
22	160	7.2	533	2 T27589	probable glycosyl
23	158	7.1	414	2 D70723	probable glycosyl
24	154.5	6.9	391	1 A69784	macrolide glycosyl
25	149	6.7	428	2 C70670	flavonol 3-O-gluc
26	148.5	6.7	471	1 S08325	probable glycosyl
27	147.5	6.6	395	2 AB0242	flavonol 3-O-gluc
28	147.5	6.6	530	2 JCS656	UDP glucuronosyltr
29	146.5	6.6	530	2 S17512	glucuronosyltransf

30	146	6.5	745	2 T23893	hypothetical prote
31	145.5	6.5	471	2 S01037	flavonol 3-O-gluc
32	143.5	6.4	471	2 S01052	flavonol 3-O-gluc
33	139.5	6.3	527	2 S15089	glucuronosyltransf
34	138.5	6.2	398	2 A97146	probable glycosyl
35	138.5	6.2	530	2 A48633	glucuronosyltransf
36	138	6.2	531	2 A35343	glucuronosyltransf
37	137.5	6.2	408	2 T44860	glycosyltransferase
38	137.5	6.2	533	2 T34458	hypothetical prote
39	137	6.1	533	2 A39092	glucuronosyltransf
40	136	6.1	382	2 T46519	probable glycosyl
41	135	6.1	535	2 T24309	hypothetical prote
42	135	6.1	535	2 T27582	hypothetical prote
43	133.5	6.0	361	2 P87565	hypothetical prote
44	133.5	6.0	530	2 S07390	glucuronosyltransf
45	133	6.0	531	2 T33980	hypothetical prote

ALIGNMENTS

RESULT 1

S52583
ctxX protein - Erwinia herbicola
C:Species: Erwinia herbicola
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999
C:Accession: S52583
R:Lin, Y.P.; Lai, B.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.
Mol. Gen. Genet. 245, 417-423, 1994
A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.
A:Reference number: S52583; MUID:95107237; PMID:7808390
A:Accession: S52583
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-431 <LIN>
A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21261.1; PID:G148395
C:Superfamily: glycosyltransferase

Query Match 80.4%; Score 1794; DB 2; Length 431;

Best Local Similarity 82.4%; Pred. No. 3.1e-132;

Matches 355; Conservative 19; Mismatches 57; Indels 0; Gaps 0;

Qy	1	MSHFAVIAPPPFHVRLQNLAQELVARGHRTVFQQHDKALVTGSDIGFTVGLQTHP	60
Db	1	MSHFAAIAPFFYSHVRLLENLAQELVARGHQTFTFQQQDTKPLIQHPHIGFHAVGESSHP	60
Qy	61	PGSLSHLLHAAHPGLGPMRLINEMARTSDMLCRELPAAFEALQIEGVIVDQMEPAGAV	120
Db	61	SGSLSLHLLHAAHPGLGPMRLINEMARTSDMLCLALPFAFNDLADGVIVDQMEPAGAL	120
Qy	121	VAPASGLPFVSVACALPLNRPGLPLAVMPPEYGTSDAARERTTSEKIYDWLVRHDRV	180
Db	121	VAEALGLPFVSVACALPLNRPGLPLAVMPPEYGTSDAARERTTSEKIYDWLVRHDRV	180
Qy	181	IAHACEMGLAPREKLIHCFSPLAQISOLIPELDPPKALPDCFFHVGPLRQPGTGPSS	240
Db	181	IAHSHRMGLAPRDKLHQCFSPLAQISOLIPELDPPKALPACFFHVGPLRQTSTSS	240
Qy	241	TSYFSPDKPRIFASLGTQLQHYGLFRTIAKACEVDQAULLAHCGGLSATQAGELARG	300
Db	241	PLYFSPDKPRIFASLGTQLQHYGLFRTIAKACEVDQAULLAHCGGLSATQAGELARG	300
Qy	301	GDIOVDFDQSNALSOAQTLTHGGMNTVLDAISRTPLLALPLAFDPQGVASRVVHG	360
Db	301	RFTQVVDFAQSAALSOAQTLTHGGMNTVLDAISRTPLLALPLAFDPQGVASRVVHG	360
Qy	361	IGKRASRFTTSHALAQIRSLTNTDYPQMTKIQAALFACGTTFAADIVBQAMRTQCP	420
Db	361	IGKRASRFTTSHALAHMRAALLTNDYDPMRTKIQTALRLAGGTTAAADIVBQAMRTQCP	420
Qy	421	VLGGQDYATAL 431	
Db	421	VLXGGYATAL 431	

```

RESULT 2
37802
rtx protein - Erwinia uredovora
Species: Erwinia uredovora
Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
Accession: B37802
Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yanano, S.; Izawa, Y.; Nakamura, K.; Harashim
Bacteriol. 172, 6704-6712, 1990
Title: Blucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio
Reference number: A37802; MUID:9107214; PMID:2254247
Accession: B37802
Status: preliminary
Molecule type: DNA
Residues: 1-431 <MIS>
Cross-references: GB:D90087; NID:g216681; PIDN:BA14125.1; PID:g216683
Superfamily: glycosyltransferase

Query Match      80.2%; Score 1790; DB 2; Length 431;
Best Local Similarity 80.5%; Pred. No. 6.3e-132;
Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;

y 1 MSHFAVIAPPPSHVRALQNLAEVARGHRVTFPPQHDCKALVTGSDIGFTVGLQTHP 60
b 1 MSHFAVIAPPPSHVRALQNLAEVARGHRVTFPPQHDCKALVTGSDIGFTVGLQTHP 60
y 61 PGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVDQMEPAGV 120
b 61 PGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVDQMEPAGV 120
y 121 VAPASGLPVSVACALPLNREBGLPLAVMPPEYGTSDAARERTTSKIIDWLRRHDRV 180
b 121 VAPASGLPVSVACALPLNREBGLPLAVMPPEYGTSDAARERTTSKIIDWLRRHDRV 180
y 181 IAHACRMGLAPREKLHCFSPALQISQLPELPFPRKALPDCFHAVGPLRQPGTSGS 240
b 181 IAHACRMGLAPREKLHCFSPALQISQLPELPFPRKALPDCFHAVGPLRQPGTSGS 240
y 241 TSYPSPDKPRIPASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGLSATQAGELARG 300
b 241 TSYPSPDKPRIPASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGLSATQAGELARG 300
y 301 GDIQVDFADQAAALSOAQLTTHGGMVLDALASRTPLALPLAFDQPGVARSRIYHG 360
b 301 GDIQVDFADQAAALSOAQLTTHGGMVLDALASRTPLALPLAFDQPGVARSRIYHG 360
y 361 IGRASRFTTSALARQRTSLTNTDYPQRTKIOALRLAGGTPAADIIVEQAMRTCPQ 420
b 361 IGRASRFTTSALARQRTSLTNTDYPQRTKIOALRLAGGTPAADIIVEQAMRTCPQ 420
y 421 VLSGGQDYATAT 431
b 421 VLSGGQDYATAT 431

RESULT 3
552980
zeaxanthin glucosyl transferase - Erwinia herbicola
Species: Erwinia herbicola
Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999
Accession: S52980
Hundle, B.; Alberti, M.; Nievelstein, V.; Beyer, P.; Kleinig, H.; Armstrong, G.A.; Bur
Mol. Gen. Genet. 245, 406-416, 1994
Title: Functional assignment of Erwinia herbicola Eho 10 carotenoid genes expressed in
Reference number: S52976; MUID:95107236; PMID:7808389
Accession: S52980
Status: preliminary
Molecule type: DNA
Residues: 1-413 <HUN>
Cross-references: ENEL:M87280; NID:g148404; PIDN:AAA64979.1; PID:g148410
Superfamily: glycosyltransferase

```

```

Query Match      45.7%; Score 1018.5; DB 2; Length 413;
Best Local Similarity 48.8%; Pred. No. 8.8e-72;
Matches 211; Conservative 62; Mismatches 128; Indels 31; Gaps 5;

Qy 1 MSHFAVIAPPPSHVRALQNLAEVARGHRVTFPPQHDCKALVTGSDIGFTVGLQ 57
Db 1 MSHFAVIAPPPSHVRALQNLAEVARGHRVTFPPQHDCKALVTGSDIGFTVGLQ 57
Qy 58 THP-PGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVDQMEP 116
Db 58 THP-PGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVDQMEP 116
Qy 52 FYPLPASVQQAQRNVQOSQWENLLRLIAAMSSLDVLCQQLPAILQRLAVDALIVDEMEP 111
Db 52 FYPLPASVQQAQRNVQOSQWENLLRLIAAMSSLDVLCQQLPAILQRLAVDALIVDEMEP 111
Qy 117 ACANVAEASGLPVSVACALPLNREBGLPLAVMPPEYGTSDAARERTTSKIIDWL 176
Db 117 ACANVAEASGLPVSVACALPLNREBGLPLAVMPPEYGTSDAARERTTSKIIDWL 176
Qy 112 AGSLVSAALGPPFISIAACALPVNREBGLPLAVMPPEYGTSDAARERTTSKIIDWL 171
Db 112 AGSLVSAALGPPFISIAACALPVNREBGLPLAVMPPEYGTSDAARERTTSKIIDWL 171
Qy 177 HRYVIAHACRMGLAPREKLHCFSPALQISQLPELPFPRKALPDCFHAVGPLRQPGT 236
Db 177 HRYVIAHACRMGLAPREKLHCFSPALQISQLPELPFPRKALPDCFHAVGPLRQPGT 236
Qy 172 HGQTLRHAQRFGLPERRLOECLSPALQISQSVLPALDFERRALPNCFFVVGALR 226
Db 172 HGQTLRHAQRFGLPERRLOECLSPALQISQSVLPALDFERRALPNCFFVVGALR 226
Qy 237 PGSSTSYFESP-----DKPRIPASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGL 289
Db 237 PGSSTSYFESP-----DKPRIPASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGL 289
Qy 227 -----YQPPQVSPSPRSTPRIFASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGL 280
Db 227 -----YQPPQVSPSPRSTPRIFASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGL 280
Qy 290 SATQAGELARGGDIQVDFADQAAALSOAQLTTHGGMVLDALASRTPLALPLAFDQ 349
Db 290 SATQAGELARGGDIQVDFADQAAALSOAQLTTHGGMVLDALASRTPLALPLAFDQ 349
Qy 281 TPAQADSLVACGATEVVSFVDPQRYVAEANLVTHGGMVLDALASRTPLALPLAFDQ 340
Db 281 TPAQADSLVACGATEVVSFVDPQRYVAEANLVTHGGMVLDALASRTPLALPLAFDQ 340
Qy 350 PGVARSRIYHGIKGRASRFTTSALARQRTSLTNTDYPQRTKIOALRLAGGTPAADI 409
Db 350 PGVARSRIYHGIKGRASRFTTSALARQRTSLTNTDYPQRTKIOALRLAGGTPAADI 409
Qy 341 PAVARLVYVNGLRGVRFRARQQLTDAEIAQLGLDTHLQRIATARQQLNDAGTTPRAAT 400
Db 341 PAVARLVYVNGLRGVRFRARQQLTDAEIAQLGLDTHLQRIATARQQLNDAGTTPRAAT 400
Qy 410 IVEQAMRTCPQV 421
Db 410 IVEQAMRTCPQV 421
Qy 401 LIEQAIAGESV 412
Db 401 LIEQAIAGESV 412

RESULT 4
574500
zeaxanthin glucosyl transferase crtX - Synecocystis sp. (strain PCC 6803)
NAlternate names: hypothetical protein slr1125
Species: Synecocystis sp.
Accession: PCC 6803
Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
Variety: PCC 6803
R.Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
Reference number: S74322; MUID:97061201; PMID:8905231
Accession: S74500
Status: nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-419 <KAN>
Cross-references: ENEL:D90899; GB:AB001339; NID:g1651550; PIDN:BA16652.1; PID:g16517;
Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
Superfamily: glycosyltransferase

Query Match      25.0%; Score 558.5; DB 1; Length 419;
Best Local Similarity 31.7%; Pred. No. 6.8e-36;
Matches 132; Conservative 91; Mismatches 196; Indels 7; Gaps 5;

Qy 1 MSHFAVIAPPPSHVRALQNLAEVARGHRVTFPPQHDCKALVTGSDIGFTVGLQTHP 60
Db 1 MSHFAVIAPPPSHVRALQNLAEVARGHRVTFPPQHDCKALVTGSDIGFTVGLQTHP 60
Qy 61 PGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVDQMEPAGV 119
Db 61 PGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVDQMEPAGV 119
Qy 121 VAPASGLPVSVACALPLNREBGLPLAVMPPEYGTSDAARERTTSKIIDWLRRHDRV 180
Db 121 VAPASGLPVSVACALPLNREBGLPLAVMPPEYGTSDAARERTTSKIIDWLRRHDRV 180
Qy 181 IAHACRMGLAPREKLHCFSPALQISQLPELPFPRKALPDCFHAVGPLRQPGTSGS 240
Db 181 IAHACRMGLAPREKLHCFSPALQISQLPELPFPRKALPDCFHAVGPLRQPGTSGS 240
Qy 241 TSYPSPDKPRIPASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGLSATQAGELARG 300
Db 241 TSYPSPDKPRIPASIGTLQGHRYGLFTIAKAEVDQAQLLAHCGGLSATQAGELARG 300
Qy 301 GDIQVDFADQAAALSOAQLTTHGGMVLDALASRTPLALPLAFDQPGVARSRIYHG 360
Db 301 GDIQVDFADQAAALSOAQLTTHGGMVLDALASRTPLALPLAFDQPGVARSRIYHG 360
Qy 361 IGRASRFTTSALARQRTSLTNTDYPQRTKIOALRLAGGTPAADIIVEQAMRTCPQ 420
Db 361 IGRASRFTTSALARQRTSLTNTDYPQRTKIOALRLAGGTPAADIIVEQAMRTCPQ 420
Qy 421 VLSGGQDYATAT 431
Db 421 VLSGGQDYATAT 431

```

120 VVHAEAGSLFVSVACALPLNREPGCLPLAVMPFFYGTSDAARERTTTSEKIYDMLRRHRD 179
121 TIGRLGIPFISCSAALNREPTTPYATFWPYPDSWLGQNLRLGGLNRAKPTITA 180
180 VIAHACRMGLAPREKLNHCFSPLAQISQLIPELDPFKALPDCFHAGVPLRQPGTSGS 239
181 LINDYRQWNLPAQSSPNDRYSPLAQISQPAAPFPRFECPLPSRFHTGFFHSNVCGRDIA 240
240 STSYFPPGDKPRIFASLGTGLOHRYGLFRTIAKACEVDQAQLLHAHCGGLSATQAGSLAR 299
241 DFPWEQUTDOPITIIYASLGTIONQLMSTFKTIAACMDLDAQLIIS-LGGAKLESMPALP- 298
300 GGDIOVVDFADQSAALSOAQLTIHFGWNTVLDIAASRTPLLALPLAFDQPGVASRIYVH 359
299 -GNPLVWVYAPOLELLORTALTITIHAGINTTLECLNNAVPMVAIPILANDQGVARIATA 357
360 GIGK-RASRFTTSHALARQIRSLTNTDYPQRMTKIOAALRLAGGTTPAAADIVEQ 413
358 GVGEFIPLSKLNTNN-LRAALEKVLITSDSVKRNLTQLQAIAKTAGGLTKAADIEQ 412

SULT 5
7104
Probable glycosyl transferase - Streptomyces coelicolor
Species: Streptomyces coelicolor
Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
Accession: T37104
Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
Submitted to the EMBL Data Library, August 1999
Reference number: Z21588
Accession: T37104
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-407 <SAU>
Cross-references: EMBL:AL109950; PIDN: CABS2955.1; GSPDB: GN00070; SCOEDB: SCJ4.21
Experimental source: strain A3(2)
Genetics:
Gene: SCOEDB: SCJ4.21
Superfamily: glycosyltransferase

Query Match 11.7%; Score 262; DB 2; Length 407;
Best Local Similarity 26.8%; Pred. No. 8.7e-13;
Matches 123; Conservative 59; Mismatches 165; Indels 112; Gaps 18;

1 MSHPAVTAPPPFHVRLQNLQAQLVARGHRVTF--FQCHDCKALVTCSDIGFTVGLQT 58
1 MSTLAFTLNIGHGHNPLVAELVARGHTVYHTTFAFREEIAATGANV-----RL 53
59 HPPCSLSHLHLAHLPL--GPSMLRLNEMARTSDMLCRELPAAFHAIQTEG---VIVDQ 113
54 YPGGD-----QPLDPDPAPITLMEALARTSLDL---LPVLAELDRDRDPLIVHDS 101
114 MEPAGVVAEAGSLFPFVSVACALPLNREPGLPLAVMPPEYGTSDAARERTTSEKIYDML 173
102 ACPWGALAAVLGLVLPVAVSSFTTFAYNRHVPSF-----TRASRE----- 139
174 MRHRDRTAHACRMGLAPRKLHHCFSPLAQISQL-----IPELDFPRKALPDCFHAGVP 229
140 -----LLAGAAAR--PRNLAVGVGARLARLRRAATGVPVLD-----LAD 177
230 LRQPGTSGSSTVSFPPS-----PDKPR-----IFASLIGTLQG 261
178 IRQPLNLVYTSRAFQPAVEEFTSRFVGPISGARPPDSPFNRLRDPVLYASLIGTVFN 237
262 HRYGLFRTIAKACEVDQAQLLHAHCGGLSATQAGELARGSDIQVWDFADQSAALSOAQLT 321
238 ADPLLLRTFATALGPL-AGTVVWSGTQDPAALGELP--GNVLARRSVPOLEVLDRRAALP 294
322 ITHGWNVTVLDIASRTEPLLALPLAFDQPGVASRIYVHGICK--RASRFT--TSHALAQ 377
295 ITHGWNVNEALFAGVPELLLVPGADQPMWARRVVELGAGLSIRTENITSDSVARAR 354
378 IRSLLTNTDYPQRMTKIOAALRLAGGTTPAAADIVEQAMR 416

[illegible]

RESULT 7

T5110
glycoyltransferase (EC 2.4.1.-) clef [validated] - Streptomyces antibioticus (ATCC 11899)
C:Species: Streptomyces antibioticus
A:Variety: ATCC 11899
C:date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T5110
R:Quiros, L.M.; Aguirrezabalaga, I.; Olano, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z25300
A:Accession: T5110
A:Status: preliminary; translated from GB/EMBL/DBJ

Molecule type: DNA
A:Residues: 1-424 <QOI>
A:Cross-references: EMBL:AF055579; PIDN:AA012648.1
A:Experimental source: ATCC 11891
C:Genetics:
A:Gene: ole1
C:Function:
Description: Intracellular inactivation of oleandomycin by glycosylation [validated, M]
Note: very specific for oleandomycin
Keywords: glycosyltransferase; hexosyltransferase

Query Match 9.7%; Score 217; DB 2; Length 424;
Best Local Similarity 24.7%; Pred. No. 3e-09;
Matches 111; Conservative 69; Mismatches 185; Indels 84; Gaps 19;

3 HFAVIAPFFSHVRAALONLAQELVARGHRTVFQOQDCKALV-----TGSDIGFQTVGLQ 57
14 HIFPNIPGHCHVNPVSLGIQVGLVARGVSYGIDTFPGAQVKGAGRTAVVYGF----- 67
58 THPPGSLHLLHLLAHPGLPSMLRLINEMARTSDMLC---RELPAAFHALQIEG---VIV 111
68 -----ILPEEFNPEELAEEDQGRWACSLAEAFRLPQLRTATPTTGRDLIVY 115
112 DQEPAGAVVAASGLPFV-----SVACA-----LPLNREPGLPL---AVMPPEYGTSDA 158
116 DIASWPAPVLGRKWDIPFVQLSPTSVAVEGFEDVPAVQDPTADRGEEAAP--AGTGA 173
159 AREYTTSEKIYDLMRHRDRIA---HH-----ACRMGLAPREKLUHCFSPLAQISQLI 210
174 --BEGAAE---DGLVRFTRLSAFLEEHGVDTFATEFLIAPNRCIVGCRAP-----SQIK 224
211 PELDFPRKALPOCFHAGPL--RQPGT---PGSSTSYFPSPDKPRIFASLTGLQGHRY 264
225 GD-----TVGDNFTFVGYGDRSHOGTWEFGHG-----RVLIIALGSAFTDHL 270
265 GLFTTAKACEVDAQULLAHCGLSATQAGELARGDGIQVDFADQSAALSQAQTLTIH 324
271 DFYRTCLSAVDGLDWHVLSVGRFVDPADLGEVP--PNVEVHWQVPLDILTKASAFITH 328
325 GGMNTVLDATASRTPLALPLADQPVASRYVYHIGKRAER-FTTSHALARQISLIT 383
329 AGMGTWEALSNVAVPVAQVQIAEQTNWAEIYELGLGRHPRDQVTAELREAVLAVAS 388
384 NTDVPORMKIOALRLAGGTAAADIVE 412
389 DPGVAERLAARQIREAGGARAADILE 417

RESULT 8
369851
acrolide glycosyltransferase homolog YJ1C - Bacillus subtilis
Species: Bacillus subtilis
Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
Accession: C69851
Authors: F. Ogasawara, N. Moszer, I. Albertini, A.M. Alloni, G. Azevedo, V. Berter
Kunst, P. Brogi, S. Brouillet, S. Bruschi, C.V. Caldwell, B. Capuano, V. Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Etian, K.D.; Etrington, J.; Fabret, C.; Ferrari, E.
ature 390, 249-256, 1997
Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
ech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Coetter, P.; Koningsstein, G.; Krosh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Maueel
r, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstira, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
Authors: Yoshikawa, H.P.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
Reference number: A69580; MUID:98044033; PMID:9384377
Accession: C69851
Status: Preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA

A:Residues: 1-392 <KUN>
A:Cross-references: GB:Z99110; GB:AL009126; NID:G2633472; PIDN:CAB13079.1; PID:G2633576
A:Experimental source: strain 168
C:Genetics:
A:Gene: yj1C
C:Superfamily: glycosyltransferase

Query Match 9.1%; Score 204; DB 1; Length 392;
Best Local Similarity 19.8%; Pred. No. 2.8e-08;
Matches 84; Conservative 82; Mismatches 206; Indels 52; Gaps 9;

3 HFAVIAPFFSHVRAALONLAQELVARGHRTVF-----FOHDCALVTGSDIGPQT 53
5 HISMNIPATGHVNPVTLALVEKCEKGRVTVATTEFPAPVQQAQGEALIVTSLNID- 63
54 VGLQTHPPGSLHLLHLLAHPGLPSMLRLINEMARTSDMLCRELPAAFPALQIEGVIVQ 113
64 -----PKQIREMMEKNDAPL-----SLKESLSILPOLELYKDDQDPLIYDF 107
114 MEPAGAVVAASGLPFVSVACALPLNREPGLPLAVMPPEYGTSDAARERTTSEKIYDML 173
108 VALAGKLFABKLVAVPVTKJCSYVAQNES-----FOLGNEMLKKIREAEAEFKAYL 158
174 MRHRDRIAHHACRMGLAPREKLUHCFSPLAQISQLIPELDFPRKALPOCFHAGVSLROP 233
159 EQEKLPAVSFEQ---LAVPEALNIVMP-----KSFQIQHETEDDRFCFVGPSSLGE 206
234 QGTFPGSSTSYFPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDAQULLAHCGLSATQ 293
207 R--KEKESLIIDKDRPLMLISUGTAFNAWPEPKYKCIKAFROSSWQVMSVGTIDPES 264
294 AGEIARGDGIQVDFADQSAALSQAQTLTIHGMNTVLDATASRTPLALPLADQPGVA 353
265 LEDIP--ANTIRQSVLPQLEVEKADLFISHGGMNSTMEAMNAGVPLVVPVQYEQELTA 322
354 SRIVYHIGKRAER-FTTSHALARQISLITNTDYQRMKIOALRLAGGTAAADIVE 412
323 NRVDGLGLVLPKEEYTVSSQEAQVSSQDQELLSRVKNMQKQVKEAGGAERAAEIE 382
413 QAMR 416
383 AFMK 386

RESULT 9
F75587
Probable glycosyltransferase - Deinococcus radiodurans (strain R1)
Species: Deinococcus radiodurans
Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
Accession: F75587
Authors: O.; Eisen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
White, O.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Me
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
Accession: F75587
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:G460670; PIDN:AAF12451.1; PID:G646074
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0329
A:Map position: 2
C:Superfamily: glycosyltransferase

Query Match 8.9%; Score 198.5; DB 2; Length 418;
Best Local Similarity 22.9%; Pred. No. 8.1e-08;
Matches 98; Conservative 79; Mismatches 210; Indels 41; Gaps 15;

6 VIAPFFSHVRAALONLAQELVARGHRTVFQOQDCKALVTGSDI---GFQTVGLQIEPPG 62

5 IASQPIAGHVPLREIARELARRGHELRWYTCRKFQPOVEAAGAVWEGF--VHARDYDDA 62

63 SLSHLLHLAAHPLGSMRLINEMARTSOMLCRELPAAPHAL-----QIEGVIVOME 115

63 NFD-----AAPGERSARRGLAQLLFDLRHVFYQMGQGYDLRLERARWRDPVVLADQTV 117

116 PAGAVVAASGLPFFSVACALPL---NREP---GLPLAVMPFEYGTSDAARER--YTTSE 167

118 GA-ALLREELGPPVALLGLVPLGFIASRDYAPFGGLAPLG---GAQQLNRALQVLTE 173

168 KIYDMLRRHRVIAHACRMGLAPREKLHCFSPLOAISOLIPELDPFRKALPDCPHAV 227

174 RV---VFRGISQLADLCTWGL-PAASFAPVAPSLMQLPSPVERLEYLRDLPPQVRFI 229

228 GPLROPOG--TPGSSTVSPSPDKPRIFASLGTQLGHRVGLFRTIAKACEVDQAQLLAH 285

230 GLPLVPAGEFTPAWNDILTRFVVLVTOGLATREQLRLPALMALADEVDLVVAA- 288

286 CGLSATQAGELARGDQIOWDFADQSAALSOAQLTIFHGMNTVLDIASRTPLALPL 345

289 --GVAPERLGLPL--ANARAAAFVFTALLPRHVASLNGYGGTQVQALSHGVPCVVAGR 344

346 AFDPQGVASRIYVHIGKXRSFTTSHALAR-QIRSLTNTDYQRMTKIOAALRAGGT 404

345 SEDKAEVAAVAYSAGNLGTATPSPARIRAAVRVUREPAFREQAGHGLHELRQHDAP 404

405 PAAADIVE 412

405 READWLE 412

RESULT 10

5906

acridole glycosyltransferase homolog yojK - Bacillus subtilis

Species: Bacillus subtilis

Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

Accession: H69906

Kunst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Arevedo, V.; Berton, S.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chodura, E.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Fritschy, J.N.; Fujita, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallet, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Jetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadaie, Y.; Sato, T.; Scanlon, J.; Schleicher, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron, J.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tognoni, K.; Tosato, V.; Uchiyama, K.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Zumbach, E.; Zumbach, E.; Zumbach, E.

Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

Reference number: A69580; MUID:98044033; PMID:9384377

Accession: H69906

Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA

Residues: 1-287 <KUN>

Cross-references: GB:Z99114; GB:AL009126; NID:G2634230; PIDN:CAB13834.1; PID:G2634335

Experimental source: strain 168

Genetics:

Gene: yojK

Superfamily: glycosyltransferase

Query Match 8.8%; Score 135.5; DB 2; Length 287;

Best Local Similarity 26.5%; Pred. No. 8.5e-08;

Matches 44; Conservative 46; Mismatches 73; Indels 3; Gaps 2;

252 IPASLGTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARGDQIOWDFADQ 311

116 LFSMGITFNKQFNQECVCKDFDKVWLSIGKIKTSELNDIPE--NFIVRPVVPQ 173

312 SAALSOAQLTIFHGMNTVLDIASRTPLALPLAFDPQGVASRIYVHIGKXRSRTTS 371

174 LEILKASLFTVTHGMNSTSEGLYFETPLVLPIMGDQFVWADQVEKVGAGKVINKEELS 233

372 HALARQ-IRSLTNTDYQRMTKIOAALRAGGTAAADIIVEQAMR 416

234 ESLKKEITQEVNNSRYAKAKXBIGQSLKAAAGSKCAADSILEAVK 279

RESULT 11

JS0636

glycosyl transferase (EC 2.4.1.1-) - Streptomyces lividans

Species: Streptomyces lividans

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999

C/Accession: JS0636

R/Jenkins, G.; Cundliffe, E.

Gene 103, 55-62, 1991

A>Title: Cloning and characterization of two genes from Streptomyces lividans that confere antibiotic resistance to streptomycin

A/Reference number: JS0635; MUID:92104506; PMID:1761231

A/Accession: JS0636

A/Molecule type: DNA

A/Residues: 1-418 <JEN>

A/Cross-references: GB:M74717; NID:G153345; PIDN:AAA26780.1; PID:G153347

A/Experimental source: strain TK21

C/Comment: This enzyme transfers glucose from UDP-glucose to macrolide antibiotics.

C/Genetics:

A/Gene: mgt

C/Superfamily: glycosyltransferase

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 8.1%; Score 181.5; DB 2; Length 418;

Best Local Similarity 22.7%; Pred. No. 1.7e-06;

Matches 101; Conservative 67; Mismatches 186; Indels 91; Gaps 17;

2 SHFAVIAPPFFSHVRALQNLQALVARGHRTVTFQOHDCKKALVTGSDIGFTQVGLQTHPP 61

23 AHIAFMSIALHGHVNPSEVIRELVARGHRTVTFQOHDCKKALVTGSDIGFTQVGLQTHPP 80

62 GSLSHLHLAAMP--LGPSMLRLINEMARTSDMLCRLPAAFPALQIEGVIVDQMEPAGA 119

81 GP-----DADPFWGSLTLDNVFPFLADALQSLPQALQAYEGDEPLVLEHDIASYAR 133

120 VVAASGLPFFSVACALP---LNEPGLPLAVMPFEYGTSDAARERTTSEKIYDMLAR 175

134 VLGRWEVFISSLSPQVAVWEGYEGVEGPMWEPKTERGQAVYARE-----HAML-- 185

176 RHDVIAHACRMGLAPREKLHCFSPLOAISOLIPELDPFRKALPDCPHAVGLPQPOG 235

186 EENGITDHPDPFGIPDRSLV-----LIP-----KAL-----QPHA 216

236 TPGSSTVSP-----PSPDKPRIFASLGTQGHRYGLFRTIAKACEVD 278

217 DRVDETITTFVGCACQGRDAGDWARPSEGAKEKVLVLSGSAFTKQPAFYRECVRAGELP 276

279 AQLLAHCGGLSATQAG---ELARGCD---IQWDFADQSAALSOAQLTIFHGMNTV 331

277 -----GMHTVLQVGRHVDPAELGDVDPNVVETWVPQLAILQQADLFVTHAGAGSQ 328

332 DALASRTPLALPLAFDPQGVASRIYVHIGKXRSRTTSHALARQIR-----SLITNTDY 387

329 EGLATATPMIAVPOADQFQGNADML--QGLGV-ARTLPTTEATAKALRTAALALVDDEV 385

388 PORMTKIOAALRAGGTAAADIIVE 412

386 AARLKEIQARMAQEAQTRGPADLIE 410

RESULT 12

A46292

zeaxanthin glucosyltransferase - Erwinia herbicola (fragment)

Species: Erwinia herbicola

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Aug-1998

C/Accession: A46292

R/Hundle, B.S.; O'Brien, D.A.; Alberti, M.; Beyer, P.; Hearst, J.E.

Proc. Natl. Acad. Sci. U.S.A. 89, 9321-9325, 1992

;;Title: Functional expression of zeaxanthin glucosyltransferase from *Erwinia herbicola*
;;Reference number: A46292; MUID:93028456; PMID:1409639
;;Accession: A46292
;;Status: preliminary; not compared with conceptual translation
;;Molecule type: nucleic acid
;;Residues: 1-54 <HUN>
;;Note: sequence extracted from NCBI backbone (NCBIP:115898)
;;Superfamily: glucosyltransferase

Query Match 8.1%; Score 180; DB 2; Length 54;
Best Local Similarity 61.1%; Pred. No. 1.6e-07;
Matches 33; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

y 306 VFDQSAALSQAQLAITHGGMNTVLDIASRTPLALPLAFDQGVASRIYVH 359
b 1 VSFVDQPVYAEANLVITHGGINTVLDALAAATVLAPELSPDQPAVARLVYN 54

RESULT 13
C5423
-hydroxyacylphingosine 1-beta-glucosyltransferase (EC 2.4.1.45) - human
;Alternate names: UDP-galactose:ceramide galactosyltransferase
;Species: Homo sapiens (man)
;Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 26-Aug-1999
;Accession: J05423
;Kapitonov, D.; Yu, R.K.
;Biochem. Biophys. Res. Commun. 232, 449-453, 1997
;Title: Cloning, characterization, and expression of human ceramide galactosyltransferase
;Reference number: J05423; MUID:97242209; PMID:9125199
;Accession: J05423
;Molecule type: mRNA
;Residues: 1-541 <KAP>
;Cross-references: GB:U62899
;Experimental source: fetal glioma cell
;Comment: This enzyme catalyzes the final step of galactosylceramide synthesis.
;Genetics:
;Gene: cgt
;Superfamily: glucuronosyltransferase
;Keywords: Glycoprotein; glucosyltransferase; hexosyltransferase
;472-492/Domain: hydrophobic #status predicted <HYD>
;538-540/Region: endoplasmic reticulum retention signal #status atypical
;78,333,442/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 7.9%; Score 175.5; DB 2; Length 541;
Best Local Similarity 20.1%; Pred. No. 7e-06;
Matches 96; Conservative 71; Mismatches 173; Indels 137; Gaps 19;

y 6 VIAPP--FFSHVRALQNLAEVARGHRTFFQHDCKALVTGSDIG-FQTVGLQTHP-- 60
b 24 IIVPPIFESHMTVEKTLASALHERGHTVFL-----LSEGRDIAPSNHYSLQRYPGI 76
y 61 -----PGSLSHLLHLAAFLPLGPMRLINEMARTSDMLCRELPAAFH 102
b 77 FNSTSDAFLQSKMRNIFSGRLTAI-----ELFDLIDHYTKNDMM-----VGNH 121
y 103 ALQTEG-----VIYDQMEPAGAVAAEAGLPFVSVACALPLNPEG--LPVAMPPE 152
b 122 AL-IQGLKKEKFDLLVDPNDMCGFVIAHLGVKAVFSTGLWYPAEAGAPAPLAYVP-E 179
y 153 YGTSDAABERYTTSEKIYDMLMRHRDRIAHAACWCLA-----PREKLHCFSPQAISQ 208
b 180 FNSLLTQMN-----LQRMKNTGYILSRGLVSLVLPK-----YERIMQKN 223
y 209 LIPE-----LDFPRKALPDCPCHAVGLPQPGTQPGSSTVSPPSPD 248
b 224 LLPEKSMYDLVHGSSLMWLTCDVALEPRPTLPNIVYVGGILTKPAS-----PLPE 274
y 249 KPRIPASGLTQGHRYGLFRTIAKACEVDAQLLHAHCGLSATQAGLAR----- 299
b 275 DLQRVANGANEHGFVLSFGAGVYLSIEDIANKL-----AGALGRUPQKVIWRF 323
y 300 -----GGDIQVVDVADQSAALSQAQLT--ITHGGMNTVLDIASRTPLALPLAFDQ 349

Db 324 SGPKPKNLGNNTKLIWLPONDLLGHSKIKAFVSHGGLNSIFETMYHGVVGVIPVFGDH 383
Qy 350 PGVASRIYVHGIKGRAS-RFTTSHALARQIRISLLTNTDYPQMTKIQAALRLAGGTP 405
Db 384 YDTWTRVQAKGKGIILEWTVTEKELYBALVKVINNPESYRQRAQKLSHIIKQDPGHP 440

RESULT 14
C97234
probable glucosyl transferase from UDP-glucuronosyltransferase family CAC2716 [imported]
;Species: Clostridium acetobutylicum
;C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
;Accession: C97234
;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
;J. Bacteriol. 183, 4823-4838, 2001
;A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
;A/Reference number: A96900; MUID:21359325; PMID:21359325
;Accession: C97234
;Status: preliminary
;Molecule type: DNA
;A/Residues: 1-407 <KUR>
;A/Cross-references: GB:AB001437; PIDN:AAK80662.1; PID:g15025750; GSPDB:GN00168
;Experimental source: Clostridium acetobutylicum ATCC824
;Genetics:
;A/Gene: CAC2716

Query Match 7.8%; Score 175; DB 2; Length 407;
Best Local Similarity 20.8%; Pred. No. 5.3e-06;
Matches 89; Conservative 178; Mismatches 108; Indels 108; Gaps 14;

Qy 10 PFFSHVRALQNLAEVARGHRTFFQHDCKALVTGSDIGFQTVGLQTHPQSLSHLLH 69
Db 10 PGYCHVNPTGLVDELVKRGEVYVYCTDEFKESIEAAGAKPISYGER-----MSNFKN 63
Qy 70 LAAPLGPSMLRLINEMARTSDMLCRELPAAFHALQIEGVIVDQMEPAGAVAAEASGLPF 129
Db 64 ANKTAGNRSYADIITKIIGMTCEIEDILKIKGMKFDYVHCHSMFCAGNIIAQILKVP 123
Qy 130 VSVACALPLNRESPGLPLAVMPPEYG-----TSDAARERTTSEKIYDMLMRH 177
Db 124 ISSFVAVATKE-----VNMQYSPITEKELSNIVSRVSEAKSRIEKRYFVKMPAI 177
Qy 178 DRVIAHACR--NGLAPREKLHCFSPQAISQISOLIDELQPPKALPDCPCHAVGP--LRQP 233
Db 178 SELM--ACRGDLNIAATK-----EFVSNIEY---YDSSFKFTGPIPKK 218
Qy 234 QCTPGSSTSYEP--SPDKPIEASIGTL---QGHRYGLFRTIAKACEVDAQLLHAHC 286
Db 219 ENLD-----FPPEKLNKKVYIISLGVFNNTNSNLNIF----- 253
Qy 287 GELSATOAGELARGGDIQVY-----DFADQSAALSQAQLTTHGG 326
Db 254 -----PEAPGSDVWVMSAYNIDTSNFINPKNFIVRNVYVQTEILKYADAATHAG 305
Qy 327 MNTVLDIASRTPLALPLAFDQGVASRIYVHGIKGRASR-FTTSHALARQIRISLLTNT 385
Db 306 MNSTDLIYNNKVPFVPIGADQLYMASRAKELGAAIVLNKOKLTAKELRESYKNYVMDA 365
Qy 386 DYPQMTKIQAALRLAGGTPAAAD 409
Db 366 SYLENIEKIRDVFEAGGYKGA 389

RESULT 15
B70878
probable transferase - Mycobacterium tuberculosis (strain H37RV)
;Species: Mycobacterium tuberculosis
;C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
;Accession: B70878
;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

```

ch completed: February 29, 2004, 14:52:45
time : 14.9967 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:35:44, Search time 13.0973 Seconds
(without alignments)
1698.885 Million cell updates/sec

File: US-09-941-947A-28
Effect score: 2231
Sequence: 1 MSHRAVIAPPFSSVRALQN.....EQAMRTCPVLGGDYATAL 431

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents_AA:*

- 1: /cgn2_6/prodata/2/aaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/aaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/aaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/aaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/aaa/ACTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1790	80.2	431	1	US-07-783-705A-2
2	960	43.0	399	1	US-08-096-623A-20
3	171.5	7.7	414	1	US-09-337-913-1
4	171.5	7.7	414	2	US-08-750-524-1
5	160.5	7.2	452	4	US-08-311-731A-6
6	142.5	6.4	426	3	US-09-320-878-8
7	142.5	6.4	426	3	US-09-105-537-20
8	142.5	6.4	426	4	US-09-141-908-8
9	142.5	6.4	426	4	US-09-657-440-8
10	142.5	6.4	3782	3	US-09-105-537-4
11	142	6.4	463	4	US-08-311-731A-7
12	138.5	6.2	530	4	US-09-356-806-113
13	138	6.2	531	5	PCT-US92-00282-6
14	137.5	6.2	408	2	US-08-926-258-2
15	137.5	6.2	408	2	US-09-120-053-2
16	137.5	6.2	408	3	US-09-198-212-2
17	137.5	6.2	408	4	US-09-789-261-2
18	137.5	6.2	431	2	US-08-576-628A-54
19	133	6.0	533	5	PCT-US92-00282-3
20	132.5	5.9	408	2	US-08-924-847A-2
21	132.5	5.9	408	3	US-09-120-052-2
22	132	5.9	409	2	US-08-924-254-2
23	132	5.9	409	3	US-09-120-249-2
24	129	5.8	523	5	PCT-US92-00282-7
25	126	5.6	524	4	US-09-356-806-40
26	123	5.5	528	4	US-09-356-806-8
27	119.5	5.4	455	3	US-09-036-987A-17

28	119.5	5.4	455	3	US-09-370-700-17
29	119.5	5.4	455	4	US-09-603-207-17
30	116.5	5.2	454	4	US-09-813-918-2
31	115.5	5.2	445	4	US-09-252-931A-17629
32	115	5.2	534	5	PCT-US92-00282-4
33	114.5	5.1	407	2	US-08-926-327-2
34	114.5	5.1	407	3	US-09-119-918-2
35	112	5.0	443	1	US-08-660-765A-4
36	111	5.0	390	3	US-09-036-987A-8
37	111	5.0	390	3	US-09-370-700-8
38	111	5.0	390	4	US-09-603-207-8
39	111	5.0	1209	4	US-09-252-991A-25844
40	108.5	4.9	530	3	US-09-180-852-2
41	108.5	4.9	740	4	US-09-624-693A-15
42	108	4.8	452	4	US-09-284-768A-22
43	108	4.8	608	4	US-09-284-768A-4
44	108	4.8	657	4	US-09-284-768A-7
45	106.5	4.8	531	5	PCT-US92-00282-5

ALIGNMENTS

RESULT 1
US-07-783-705A-2
Sequence 2, Application US/07783705A
Patent No. 5429939
GENERAL INFORMATION:
APPLICANT: Misawa, No. 5429939ihiko
APPLICANT: Kobayashi, Kazuo
APPLICANT: Nakamura, Katsumi
APPLICANT: Yamano, Shigeyuki
TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
TITLE OF INVENTION: SYNTHESIS OF CAROTENOID
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ladas & Parry
STREET: 26 West 61 Street
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10023
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: N/A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07783,705A
FILING DATE: 19911023
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-103078
FILING DATE: 21-APR-1989
APPLICATION NUMBER: JP 2-53225
FILING DATE: 05-MAR-1990
APPLICATION NUMBER: US 07/519,011
FILING DATE: 19-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schwadron, Janet I.
REGISTRATION NUMBER: 33,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-708-1935
TELEFAX: 212-246-5959
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 431 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-783-705A-2

Query Match 80.2%; Score 1790; DB 1; Length 431;


```
Best Local Similarity 80.5%, Pred. No. 1.7e-178;
Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;

1 MSHFAVIAPFFSHVRALQNLQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTH? 60
1 MSHFAAIPFFSHVRALQNLQALVARGHRVTFIQYDIKHLIDSETIGFHSVGTDSHP 60
61 PGSLSHLHLAHPGLPSMLRLINEMARTSDMLCRLPAAFHALQIEGVIVDMEPAGAV 120
61 PGALTRVLHLAHPGLPSMLKLINEMARTTDMCLRELFPQFNDLAVDGVIVDMEPAGAL 120
121 VAASGLPFFSVACALPLNPEGLPLAVMPEFYGTSDAARERTTTSEKIYDMLMRHDRV 180
121 VAERLGLPFFSVACALPLNPEGLPLAVMPEFYGTSDAARERTTTSEKIYDMLMRHDRV 180
181 IAHACRMGLAPREKLHCHSPALQISOLIPDLDPKPKALPDCFHAVGRLPQGTGSS 240
181 IAEHSHRMGLAPROKLGCEPLAQISOLIPDLDPKPKALPACFHAVGRLPETHAPSTSS 240
241 TSYPSPDKPRIPASLGTQCHRYGLFRTIAKACEEVDQALLAHCGGLSATQAGELARG 300
241 SRYTSEKPRIPASLGTQCHRYGLFRTIVKACEIDQGLLHAGCGRLTDSQCELAAS 300
301 GDIOVDFADQSAALSAQATITTHGGMNTVLDALASRTPLLLALPLAFDQPGVASRIYVHG 360
301 RHTQVDFADQSAALSAQALTAITHGGMNTVLDALNTVTPLLALPLAFDQPGVASRIYVHG 360
361 IGKEASFTTSHALAROIRSLTNTDYPQRTKIQALRLAGGTAAADIVEQAMRTCP 420
361 IGKEASFTTSHALARQMSLTNTVDFQQRWAKIQTRALRAGGTMAADIIIEQVMCTGP 420
421 VLSGGDYATAL 431
421 VLSGGGYATAL 431

RESULT 2
US-08-096-623A-20
; Sequence 20, Application US/08096623A
; Patent No. 5684238
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brikkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Hwei-Che B.
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
; Glycosylated Zeaxanthin in Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welsh & Katz, Ltd.
; STREET: 120 S. Riverside Plaza, 22nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,623A
; FILING DATE: 22-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,061
; FILING DATE: 09-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,921
; FILING DATE: 28-FEB-1991
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 07/562,674
; FILING DATE: 03-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/525,551
; FILING DATE: 18-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/487,613
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; REFERENCE/DOCKET NUMBER: AMO-006.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 655-1500
; TELEFAX: (312) 655-1501
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-096-623A-20

Query Match 43.08; Score 960; DB 1; Length 399;
Best Local Similarity 48.6%; Pred. No. 9.5e-32;
Matches 204; Conservative 58; Mismatches 124; Indels 34; Gaps 7;

QY 1 MSHFAVIAPFFSHVRALQNLQALVARGHRVTFQOHDCKALVTGSDIGF---QTVGLQ 57
DB 1 MSHFAIYAPPLYSHVAHALALEMAQGRHRTF-----LTGNVASLAEQETERVA 51
QY 58 THP-PGSLSHLLHLAHPGLPSMLRLINEMARTSDMLCRLPAAFHALQIEGVIVDMEP 116
DB 52 FYPLPASVQQAQRNVQOQSGNGLRLIAAAMSSLTDLVLCQQLPAILORLAVDALIVDMEP 111
QY 117 AGAVVAASGLPFFSVACALPLNPEGLPLAVMPEFYGTSDAARERTTTSEKIYDMLMR 176
DB 112 AGSIVAEALGLPFFISIALPVRNE--LPLPWFPHYAEKRAKRAKRAKRAKRAKRAKRA 169
QY 177 HDRVIAHACRMGLAPREKLHCHSPALQISOLIPDLDPKPKALPDCFHAVGRLPQGT 236
DB 170 HGQTLRPAQRFGLPERRLDCELSPLAQISQSVPALDPFRALPNCFTYVVALR----- 224
QY 237 PGSSTSVFSP-----DKPRIPASLGTQCHRYGLFRTIAKACEEVDQALLAHCGGL 289
DB 225 -----YQPPFQVSPSPRTPRIPASLGTQCHRYGLFRTIAKACEEVDQALLAHCGGL 278
QY 290 SATQAGELARGDIIQVYDFADQSAALSAQATITTHGGMNTVLDALASRTPLLLALPLAFDQ 349
DB 279 TPAQADSLYCGA-TEVVSFVDQPRYVAEANLVITHGGLNTVLDALAAATPVLAVPLSFDQ 337
QY 350 PGVASRIYVHGIGKASRFTTSHALAQIRSLTNTDYPQRTKIQALRLAGGTAAAD 409
DB 338 PAVAAIRVYNGLGRVSRFARQOTLADETIAQLLQDETLHORVATASQQLNDAGTPTCGD 397

RESULT 3
US-09-337-913-1
; Sequence 1, Application US/09337913
; Patent No. RE37206
; GENERAL INFORMATION:
; APPLICANT: Katsuhisa KOJURI, et al.
; TITLE OF INVENTION: GENE ENCODING GLYCOSYLTRANSFERASE AND ITS
; TITLE OF INVENTION: USES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Sherman and Shalloway
; STREET: 413 N. Washington Street
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.2
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/337,913
FILING DATE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/337,913
FILING DATE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,524
FILING DATE: December 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: S-2401
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-337-913-1

Query Match 7.7%; Score 171.5; DB 1; Length 414;
Best Local Similarity 21.4%; Pred. No. 2.9e-09;
Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
6 VIAPFFGHVRALQNLAEVLVARGHRTF-----34
7 VATTFGDGHVNPVQAEVMSRGHEVWYTKAPRSTVETGARHEPMRDAHDFGGMPR 66
35 ---FOQDCKALVTGSDIGFOTVGLQTHPPGSLHLLHAAHPGLSPMLRLINEMARTSD 91
67 EAPFQHAGLTCITNMIAGFRDPIE-----PAADQMTDLL-----102
92 MLCRELPAFAHALQIEGVIVDQMEPAGAVABASGLPFVSVACAL-----PLNREP 142
103 ALLEDFFA-----DVLVTDETFFGAGFVSERTGIPVAMIATSIYVFSRDTAPLGL-- 153
143 GLPLAVMPEYCTSDAARERTTSSEKIYDML---MRSHDRVIAHACRMGLAPREKLH 198
154 GLPPS-----SSRLGLRLNVLKQLTRVMDLRRAHVVD---RVGL-PRIR-KG 201
199 CFSPLAQISQL-----IPELDPFKALPDCFHAGVPLRQPGTSGSSTSYFPPSD--KPR 251
202 AFENIMRTPDLVLLGTVPSEFYPRGDMPEVRFVGPVSPAPPDFTPPAMWGELDSGRPV 261
252 IPASLGTQCHRYGLFRFRIAKACEVDQAQLLHACGGLSATOAGELARGG-----301
262 VEVTTQ-----TVANDAE-----RLLLPAIRALAEDVLVAVTTGAPLELEPMPA 306
302 DIQWVDFADQSAALQAQTLTHGGMTVLDIAASRTPLLALPLAFDPQGVASRIYVYHI 361
307 NVYVERFIPHALLPHVDAMVYNGYGVNTALAHGVPLV-VARTEEKHEVAARVWSGA 365
362 GKR-ASRFTTSHALAQIRSLTNTDYPQ 389
366 GVHLKKRRLSERDIRRAVAVL---DEPR 391

US-08-750-524-1

Query Match 7.7%; Score 171.5; DB 2; Length 414;
Best Local Similarity 21.4%; Pred. No. 2.9e-09;
Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
6 VIAPFFGHVRALQNLAEVLVARGHRTF-----34
7 VATTFGDGHVNPVQAEVMSRGHEVWYTKAPRSTVETGARHEPMRDAHDFGGMPR 66
35 ---FOQDCKALVTGSDIGFOTVGLQTHPPGSLHLLHAAHPGLSPMLRLINEMARTSD 91
67 EAPFQHAGLTCITNMIAGFRDPIE-----PAADQMTDLL-----102
92 MLCRELPAFAHALQIEGVIVDQMEPAGAVABASGLPFVSVACAL-----PLNREP 142
103 ALLEDFFA-----DVLVTDETFFGAGFVSERTGIPVAMIATSIYVFSRDTAPLGL-- 153
143 GLPLAVMPEYCTSDAARERTTSSEKIYDML---MRSHDRVIAHACRMGLAPREKLH 198
154 GLPPS-----SSRLGLRLNVLKQLTRVMDLRRAHVVD---RVGL-PRIR-KG 201
199 CFSPLAQISQL-----IPELDPFKALPDCFHAGVPLRQPGTSGSSTSYFPPSD--KPR 251
202 AFENIMRTPDLVLLGTVPSEFYPRGDMPEVRFVGPVSPAPPDFTPPAMWGELDSGRPV 261
252 IPASLGTQCHRYGLFRFRIAKACEVDQAQLLHACGGLSATOAGELARGG-----301
262 VEVTTQ-----TVANDAE-----RLLLPAIRALAEDVLVAVTTGAPLELEPMPA 306
302 DIQWVDFADQSAALQAQTLTHGGMTVLDIAASRTPLLALPLAFDPQGVASRIYVYHI 361
307 NVYVERFIPHALLPHVDAMVYNGYGVNTALAHGVPLV-VARTEEKHEVAARVWSGA 365
362 GKR-ASRFTTSHALAQIRSLTNTDYPQ 389
366 GVHLKKRRLSERDIRRAVAVL---DEPR 391

RESULT 4
US-08-750-524-1
Sequence 1, Application US/08/50524
Patent No. 5861293
GENERAL INFORMATION:
APPLICANT: Katsuhisa KOJIRI, et al.
TITLE OF INVENTION: GENE ENCODING GLYCOSYLTRANSFERASE AND ITS
TITLE OF INVENTION: USES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSER: Law Office of Sherman and Shalloway
STREET: 413 N. Washington Street
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: Dell System 210; Intel 80 286 Microprocessor
OPERATING SYSTEM: MS DOS 6.2
SOFTWARE: Word Perfect, Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,524
FILING DATE: December 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Richard A. Steinberg
REGISTRATION NUMBER: 26,588
REFERENCE/DOCKET NUMBER: S-2401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 549-2282
TELEFAX: (703) 836-0106
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-750-524-1

Query Match 7.7%; Score 171.5; DB 2; Length 414;
Best Local Similarity 21.4%; Pred. No. 2.9e-09;
Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
6 VIAPFFGHVRALQNLAEVLVARGHRTF-----34
7 VATTFGDGHVNPVQAEVMSRGHEVWYTKAPRSTVETGARHEPMRDAHDFGGMPR 66
35 ---FOQDCKALVTGSDIGFOTVGLQTHPPGSLHLLHAAHPGLSPMLRLINEMARTSD 91
67 EAPFQHAGLTCITNMIAGFRDPIE-----PAADQMTDLL-----102
92 MLCRELPAFAHALQIEGVIVDQMEPAGAVABASGLPFVSVACAL-----PLNREP 142
103 ALLEDFFA-----DVLVTDETFFGAGFVSERTGIPVAMIATSIYVFSRDTAPLGL-- 153
143 GLPLAVMPEYCTSDAARERTTSSEKIYDML---MRSHDRVIAHACRMGLAPREKLH 198
154 GLPPS-----SSRLGLRLNVLKQLTRVMDLRRAHVVD---RVGL-PRIR-KG 201
199 CFSPLAQISQL-----IPELDPFKALPDCFHAGVPLRQPGTSGSSTSYFPPSD--KPR 251
202 AFENIMRTPDLVLLGTVPSEFYPRGDMPEVRFVGPVSPAPPDFTPPAMWGELDSGRPV 261
252 IPASLGTQCHRYGLFRFRIAKACEVDQAQLLHACGGLSATOAGELARGG-----301
262 VEVTTQ-----TVANDAE-----RLLLPAIRALAEDVLVAVTTGAPLELEPMPA 306
302 DIQWVDFADQSAALQAQTLTHGGMTVLDIAASRTPLLALPLAFDPQGVASRIYVYHI 361
307 NVYVERFIPHALLPHVDAMVYNGYGVNTALAHGVPLV-VARTEEKHEVAARVWSGA 365
362 GKR-ASRFTTSHALAQIRSLTNTDYPQ 389
366 GVHLKKRRLSERDIRRAVAVL---DEPR 391

RESULT 5
US-08-311-731A-6
Sequence 6, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELEPHONE: 617/720-2441
TELEFAX: 617/720-3500
INFORMATION FOR SEQ ID NO: 6:
LENGTH: 452 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM TUBERCULOSIS

US-08-311-731A-6

Query Match

Best Local Similarity 7.2%; Score 160.5; DB 4; Length 452;

Matches 93; Conservative 67; Mismatches 154; Indels 169; Gaps 18;

9 PPFFSHVR-----ALONLAQELVARGHRTFFQOHCCKALVTGSDIGFQTVGLQTHPG 62

14 PVVASEVGRDAAISTTDPAPGLAARRRILFVAE-----AVT----- 51

63 SLSHLLH--LAAPHPLGSMRLINEMARTSDMLCRLPAALFHAL----- 104

52 LAHVVPFALAQSLDPSRYEVHVFACDPYRNQLGLPFRHAIHTIPSEFPGNLTQGR 110

105 -----QTEGVIVDQMEP-----AGAVVAEASGLPFVSVACALPLNREPGL 144

111 FYAMRTLRKVEADLRVLDSEIADPLVVGDLRISLSVSARLAGIPYIAANAY----- 162

145 PLAVMPPEYGTSDAAREYTTSEKIY-----DWMKMR 176

163 -----WSPYAQRFFPLDPVINTLFGVRLVXLLYLERPLLFALQCMPLANVRR 212

177 HD-RVIAHACRMGLAPREKLHCFSPLAQISQILPELDPKALPDCHAVCP-LRQPO 234

213 HGLSSLGWNLCRIFTDGDHLY-----ADVPMLMPTD-----LPANHEYLGVLWSPA 261

235 GTPGSSSTYPPSPDKPRIFASLGTLOGHRYGLFRTIAKACEEVDQALLAHCGGLSATOA 294

262 GKPTWDSLFT-DRPIVYATLGTSGGNL-----LQVLNALAELPVTVI 306

295 GELARGDIO-----VVPFADQSAALSOAQTIITHGQNTVLDIASRTPLALPLAF 347

307 AATAARGDLKTVPANAFVADYLPGEAAARSAVVWCNGSLTTQQALVAGVPVI----- 360

348 DQPGVASSRIVVH-----GIGKRSRFTTSHALARQIRSLTNTDYPQRMKIQ 396

361 ---GVAGNLDQHLNWEAVERAGAGVILLRTEL-KSQRVAGANVQVISRSEYPTSRATR 416

QY 397 ALR 399

Db 417 CLR 419

RESULT 6

US-09-320-878-8

; Sequence 8' Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A

; EARLIER FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908

; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139

; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880

; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 426

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-320-878-8

Query Match 6.4%; Score 142.5; DB 3; Length 426;

Best Local Similarity 23.1%; Pred. No. 3.3e-06;

Matches 109; Conservative 42; Mismatches 182; Indels 139; Gaps 24;

QY 13 SHVRLONLAQELVARGHRTFFQOHCCKALVTGSDIGFQTVGLQTHPGSLSHLLH--- 69

Db 12 THYGLVPLAWALLAGHEVYVASQALTTTIGS-----GLAAPPVGT-DHLIHEYR 63

QY 70 --LAAPHPLGSMRLINEMARTSDMLCRLPAALF-HALQIEGVI-----VDQ 113

Db 64 VRMAGEP-RPNHPALAPDEARPS-----PLDWDHALGTEAILAPFYLLANNDSMVD 115

QY 114 M-----EP---AGAVVAEASGLPFVSVACALPLNREPGLPLAVMPPEYGT 156

Db 116 LVDPARSQDPLVWEPTTYAGAAVQVGAARVLWG-----PDVWG--- 159

QY 157 DAARERTTSEKTYDMLMRHHRVIAHACRMGLAPREKLHCFSPLAQISQILPELD--- 214

Db 160 -SARKXZ-----VALDRQPPHEHEDPTAE--WLTWTLDRY 192

QY 215 ---PPRKALPDCHAVCP---LRQPGTSGSSTSYFP-----SPDKPRIFA 254

Db 193 GASPEEELLTQGF-TIDTPTPSURLDTGLFTVGMRVFVNGTSVVPDMLSEPPARPVCL 251

QY 255 SIGYTL-----QCHRYGLPRTIAKACEEVDQALLAHCGGLSATQAGELAR--GGDIQV 306

Db 252 TLGVASAEVLGSDGVSGQ---DILEALADLDIELVAT---LDASQRAEIRNYPKHTRFT 304

QY 307 DFADQSAALSOAQTIITHGQNTVLDIASRTPLALPLAFDQPGVASRIVVHIGIKRAS 366

Db 305 DFVPMHALLPSCSAIITHGGAGFTYATAVINAVPQVLMELWADPVKARVAEQAG--- 360

QY 367 RFTTSHALARQ-IRSLTUN-TDYPQRMTKIQALLRAGGTAAADIVEQAMR 416

b 361 FFLPPAELTPQAVRDVAVRLLDDPSVATAAHLREBTGDPFGAGIVPELER 412

RESULT 7

S-09-105-537-20

Sequence 20, Application US/09105537A

Patent No. 6265202

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.438US1

CURRENT APPLICATION NUMBER: US/09/105,537A

CURRENT FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 20

LENGTH: 426

TYPE: PRT

ORGANISM: Streptomyces venezuelae

S-09-105-537-20

Query Match 6.4%; Score 142.5; DB 3; Length 426;

Best Local Similarity 23.0%; Pred. No. 3.3e-06;

Matches 107; Conservative 42; Mismatches 191; Indels 125; Gaps 22;

Y 13 SHVRALQNLAEVLVARGHRTFFQCHDKALVTGSDIGFTVGLQTHPPGSLSLHLH--- 69

b 12 THYGLVPLAWALLAAGHEVRVASQPALTDTTIGS-----GLAAPPVGT-DHLIHYR 63

Y 70 --LAAHPLGSMRLINEMARTSDM-----LCRELPAALQIEGVVDQM----- 114

b 64 VRNAGEP-RPNHPAIAFDEARPEPLDWDHALGIEAILAPYFLLANNDMSVDDLVDPARS 122

Y 115 -----EP-----AGAVAAASGLPFVSVACALPLNRBGLPLAWMPPEYGTSDAARERY 163

b 123 WQDLVLEPTTYAGAAQVTAAGAAHVWG-----PDVWG--SARRKF 165

Y 164 TTSEKIYDMLRRHRDRIAHAACRMGLAPREKLHCFSPQAQISOLIPELD-----PPRK 218

b 166 -----VALDRQPPHREDPTAB--WLTWLDTRYGASPEEE 199

Y 219 ALPDCFHAVGP-----LRQPGTSGSTSYFP-----SPDKPRIFASLGLT-- 259

b 200 LUTGQP-TIDPTPPSLRLDTGLPTVMRYVPYNGTSVVPDMLSEPPARPRVCLTLGVARS 258

Y 260 -----QCHRYGLPRTIAKACEVEDAQLLHACGGLSATQAGELAR--GGDIQWDFADQSA 313

b 259 EVLGGDVSQG--DILEALADLDIELVAT----LDASQRAEIRNYPKHTRFTDFVPMHA 311

Y 314 ALSQAQLTTHGGMNTVLDASRTPLIALPLAFDPQGVASRIVVHGIGKASRFTTSHA 373

b 312 LLPSCSAIIHGGAGTYATAVINAVPQVMAELWDAPVKARAVABQAGAG----PFLPPAB 367

Y 374 LARO-IRSLTN-TDYPQMTKIQALRLAGCTPAAADIVEQAMR 416

b 368 LTPQAVRDVAVRLLDDPSVATAAHLREBTGDPFGAGIVPELER 412

RESULT 8

S-09-141-908-8

Sequence 8, Application US/09141908

Patent No. 650374;

GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary

APPLICANT: BETLACH, Melanie C.

APPLICANT: BETLACH, Mary

APPLICANT: MCDANIEL, Robert

APPLICANT: TANG, Li

TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a

; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold

; FILE REFERENCE: 300622002100

; CURRENT APPLICATION NUMBER: US/09/141,908

; CURRENT FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: PROV. 60/076,919

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: PROV. 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 426

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-141-908-8

Query Match 6.4%; Score 142.5; DB 4; Length 426;

Best Local Similarity 23.1%; Pred. No. 3.3e-06;

Matches 109; Conservative 42; Mismatches 182; Indels 139; Gaps 24;

QY 13 SHVRALQNLAEVLVARGHRTFFQCHDKALVTGSDIGFTVGLQTHPPGSLSLHLH--- 69

Db 12 THYGLVPLAWALLAAGHEVRVASQPALTDTTIGS-----GLAAPPVGT-DHLIHYR 63

QY 70 --LAAHPLGSMRLINEMARTSDMLCSLPAAF-HALOIEGVI-----VDQ 113

Db 64 VRNAGEP-RPNHPAIAFDEARPE-----PLDWDHALGIEAILAPYFLLANNDMSVDD 115

QY 114 M-----EP-----AGAVAAASGLPFVSVACALPLNRBGLPLAWMPPEYGT 156

Db 116 LVDFAFSWQPDVLEWPTTYAGAAQVTAAGAAHVWG-----PDVWG-- 159

QY 157 DAARERYTTSEKIYDMLRRHRDRIAHAACRMGLAPREKLHCFSPQAQISOLIPELD-- 214

Db 160 -SARRKF-----VALDRQPPHREDPTAB--WLTWLDTRY 192

QY 215 ---PPRKALPDCFHAVGP-----LRQPGTSGSTSYFP-----SPDKPRIFA 254

Db 193 GASFEELLTGQP-TIDPTPPSLRLDTGLPTVMRYVPYNGTSVVPDMLSEPPARPRVCL 251

QY 255 SLGTL-----QCHRYGLPRTIAKACEVEDAQLLHACGGLSATQAGELAR--GGDIQV 306

Db 252 TLGVASREVVLGGDVSQG--DILEALADLDIELVAT----LDASQRAEIRNYPKHTRFT 304

QY 307 DFADQSAALSQAQLTTHGGMNTVLDASRTPLIALPLAFDPQGVASRIVVHGIGKAS 366

Db 305 DFVPMHALLPSCSAIIHGGAGTYATAVINAVPQVMAELWDAPVKARAVABQAGAG--- 360

QY 367 RFTTSHALARO-IRSLTN-TDYPQMTKIQALRLAGCTPAAADIVEQAMR 416

Db 361 FFLPPAELTPQAVRDVAVRLLDDPSVATAAHLREBTGDPFGAGIVPELER 412

RESULT 9

US-09-657-440-8

; Sequence 8, Application US/09657440

; Patent No. 6509455

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/657,440

; CURRENT FILING DATE: 2000-09-07

; PRIOR APPLICATION NUMBER: 09/320,878

; PRIOR FILING DATE: 1999-05-27

PRIOR APPLICATION NUMBER: CIP OF 09/141,908
PRIOR FILING DATE: 1998-08-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 8
LENGTH: 426
TYPE: PRT
ORGANISM: Streptomyces venezuelae
IS-09-657-440-8

Query Match 6.4%; Score 142.5; DB 4; Length 426;
Best Local Similarity 23.1%; Pred. No. 3.3e-06;
Matches 109; Conservative 42; Mismatches 182; Indels 139; Gaps 24;

2Y 13 SHVRALQNLAEVARGHRVTFQCHDCKALVTGSDIGFQTVGLQTHPPGSLSHLLH--- 69
DB 12 THYGLVPLAWALLAGHEVRVASQPALTDITGS-----GLAAVPVGT-DHLIHEYR 63
2Y 70 --LAAHPLGSMRLINEMARTSDMLCRELPAP-HALQIEGV-----VQO 113
DB 64 VRMAGEP-RNHPAIAFDEARPEPLWDHALGIEATLAPYFYLLANNDMSVDD 115
2Y 114 M-----EP---AGAVVABASGLPFVSVACALPLNREPGLPLAVMPFFYGT 156
DB 116 LVDFARSHQPLVLEWPTTYAGAAVQTGAARVVLWG-----PDVWG-- 159
2Y 157 DAARERTTSKIYDMLRRHDRVIAHACRMGLAPREKHLHCFSPALQISQLIPELD-- 214
DB 160 -SARRKF-----VALDRQPPHEHREDPTAE--WLTWTLDRY 192
2Y 215 ---PFRKALPCFHAVGP---LRQPGTSGSSTSYEP-----SPDKPRIFA 254
DB 193 GASFEELLTQF-TIDTPPSLRLDGLPTVGMRYVYNGTSVVPDMLSEPPARPRVCL 251
2Y 255 SLGTL-----QGHRYGLFRTIAKACEEVDQAQLLAHCGLSATQAGELAR--GGDIQV 306
DB 252 TLGVSAREVLGGDVSQG---DILEALADLDIELVAT---LDASQRAEIRNYPKTRFT 304
2Y 307 DFADQSAALSOAQLTIHGGMNTVLDALASRTPLALPLAFDOPGVASRIYVHIGKRSAS 366
DB 305 DFVPMHALLPSCSAIIHGGAGTATAVINAVPQVMLAELWDAPVKARAVAEQAG----- 360
2Y 367 RFTTSHALAQ-IRSLTN-TDYPQRTMKIOAALRAGGTPAAADIVEQAMR 416
DB 361 FFLPPAELTPQAVRDVVRILDDPSVATAAHRLEETFGDPTPAGIVPELER 412

RESULT 10
US-09-105-537-4
Sequence 4, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 3782
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-105-537-4

Query Match 6.4%; Score 142.5; DB 3; Length 3782;
Best Local Similarity 23.0%; Pred. No. 0.0001;
Matches 107; Conservative 42; Mismatches 191; Indels 125; Gaps 22;

2Y 13 SHVRALQNLAEVARGHRVTFQCHDCKALVTGSDIGFQTVGLQTHPPGSLSHLLH--- 69

DB 414 THYGLVPLAWALLAGHEVRVASQPALTDITGS-----GLAAVPVGT-DHLIHEYR 465
QY 70 --LAAHPLGSMRLINEMARTSDMLCRELPAP-HALQIEGV-----VQO 114
DB 466 VRMAGEP-RNHPAIAFDEARPEPLWDHALGIEATLAPYFYLLANNDMSVDDIVDFARS 524
QY 115 -----EP---AGAVVABASGLPFVSVACALPLNREPGLPLAVMPFFYGTSDAARERY 163
DB 525 WQPDVLVLEWPTTYAGAAVQTGAARVVLWG-----PDVWG--SARRKF 567
QY 164 TTSEKIYDMLRRHDRVIAHACRMGLAPREKHLHCFSPALQISQLIPELD-----PFRK 218
DB 568 -----VALDRQPPHEHREDPTAE--WLTWTLDRYVAGSFE 601
QY 219 ALPCFHAVGP---LRQPGTSGSSTSYEP-----SPDKPRIFA 259
DB 602 LLTQF-TIDTPPSLRLDGLPTVGMRYVYNGTSVVPDMLSEPPARPRVCLTLGVSAR 660
QY 260 -----QGHRYGLFRTIAKACEEVDQAQLLAHCGLSATQAGELAR--GGDIQV 313
DB 661 EVLGGDVSQG---DILEALADLDIELVAT---LDASQRAEIRNYPKTRFTDFVPMHA 713
QY 314 ALSQAQLTIHGGMNTVLDALASRTPLALPLAFDOPGVASRIYVHIGKRSASRFTTSHA 373
DB 714 LPPCSAIIHGGAGTATAVINAVPQVMLAELWDAPVKARAVAEQAG-----FFLPPAE 769
QY 374 LARQ-IRSLTN-TDYPQRTMKIOAALRAGGTPAAADIVEQAMR 416
DB 770 LTPQAVRDVVRILDDPSVATAAHRLEETFGDPTPAGIVPELER 814

RESULT 11
US-08-311-731A-7
Sequence 7, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: MYCOBACTERIUM TUBERCULOSIS

J-08-311-731A-7

Query Match 6.4%; Score 142; DB 4; Length 463;

Best Local Similarity 21.5%; Pred. No. 4.2e-06;

Matches 98; Conservative 69; Mismatches 188; Indels 100; Gaps 22;

20 NLAQELVARGHRTFFQ-----HDCKALVTGSDIGFTVGLQTHPPGSLHLLHAHP 74

21 NAAEPVARRQRIUVGEAALAHVRFVLA RSL-----DPSRYEVHACDP 68

75 -----LGP-----SMRLINEMARTSDMLCRELPAAHALQIE---GVIV 111

69 RFNKLGLPLPPHPIHTVPSEEVLLKIAQGRLFYNTRTLEKYIAADRKILNFIAPDVV 128

112 DQMEPAGAVVAEASGLPVSACAL--PLNREPGPLAVMPFE--YGTSDAARERTISE 167

129 GENRLSLVSARLAGIPYIAANAYNSQARR--RPPLDPVPTWTFVGRPV-----SILY 182

168 KIY-----DWLMRRHD--RVIAHACRMGLAPREKLHHCFSPLAQISQLIPELD 214

183 RLYRLEIFALYCLPLNMLRRKHLSSGLWDLCPRTDGDVLY-----ADVPDLVPTYN 236

215 FPRKALPCFHVGP-LPQPGTGCSTSYSPSPDK9RIEASLGTLOGHRYGLPRTIAKA 273

237 -----LPANRYLGVPLWSPDVXKPTWMSLPT-DRPIIYATLGGSGK--NLLQVINA 288

274 CEEDVAQLLHACG--GLSATQAGELARGEDIQVDFADQSAALSQALTIHGGMNTVL 331

289 LADLPVTVIATAGRNHLKNVPANAF-----VADYLPGEAAARLAVVLCNGSGSPTQ 341

332 DAASRTPLALPLAFDPQGVASRIYVHGIG--KRASRTTSHALARQIRSLTNTDYPQ 389

342 QALAGVVIIGLPSNMDDHLMNEALERAGAVLLRTERLNT--BGVAAAVQVLSGAEFR- 399

390 RMTKQAAIRLAGGTPAAADIVQAMRTCPQVLSG 424

400 -----QAARL-----PKESDQTLPSRSTSKVRG 425

RESULT 12

J-09-356-806-113

Sequence 113, Application US/09356806

Patent No. 6586175

GENERAL INFORMATION:

APPLICANT: Penny, Laura

APPLICANT: Galvin, Margaret

APPLICANT: Miller, Andrew

APPLICANT: Reidy, Michael

TITLE OF INVENTION: Genotyping Human

TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and

FILE REFERENCE: SEQ-22PRV2

CURRENT APPLICATION NUMBER: US/09/356,806

CURRENT FILING DATE: 1999-07-20

NUMBER OF SEQ ID NOS: 164

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 113

LENGTH: 530

TYPE: PRT

ORGANISM: H. sapiens

J-09-356-806-113

Query Match 6.2%; Score 138.5; DB 4; Length 530;

Best Local Similarity 20.2%; Pred. No. 1.2e-05;

Matches 91; Conservative 75; Mismatches 193; Indels 91; Gaps 18;

6 VIAPFFHVRALQNLQALVARGHRTFFQHDCKALVTGSDIGFTVGLQTHPPGSL 65

27 LVKPTSEYSEWINKTIILBELVQRGHEVTLTS--SASTLVNASK--SSAILEVYPTS--- 80

66 HELHLAAHPLGFSMLRLIN-----EXARTSDMLCRE----- 96

81 ----LTKNDLEDLSLKILDRWIYGVSKNTFWSYFQELCWEYDYSNKLCKDAVLNKK 136
97 LPAAFHAIQIEGVIVDQKEPAGAVVAEASGLPFV-----SVACALPLNRE-----PGLPLA 147
137 LAMKLOESKDFVILADALNPGCBELAELEFNIPFLYSRFSVGYTFEKNGCGFLFPSPVVP 196
148 VMPFEYGTSDAARER-----YTTSEKIYDLMRRHDRVIAHACR-----MGLA 191
197 VVMSLSQPMIFMERIKNMIMKLYDFEFWQIYD--LKKWDQFYSEVLGRPTTLFETMGKA 254
192 PREKLHHCFSPLAQISQLIPELDPRKALPCFHVGPV--RQPGTSGSSTSYPPSPDK 249
255 ENMLRTYV-----DFFPRPPLPN--VDFVGLHCKPAKPLPKEMEEFVQSSGE 302
250 PRIFA-SLGT-LOGHRYGLFETIAKACEVEDAQLLHACGGLSATQAGELARGDDIQVVD 307
303 NGIVVFSGLSMISNWSBESANMIASALAIQIPQVLRFPDGKKPT-----LGSNTRLYK 356
308 PADQSAALS--QAQLTIHGGMNTVLDAIASRTPLALPLAFDPQGVASRIYVHGIGKRA 365
357 WLPQNDLGHPTKATFTGGTNGIYEALYHIGIPWGIPLPADQHDNIAMKAKGAALS 416
366 S-RFTTSHALARQIRSLTNTDYPQRTKI 394
417 DIRTWSSRDLNALKSVINDPVYKENVMKL 446

RESULT 13

PCT-US92-00282-6

Sequence 6, Application PC/TUS9200282

GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.

APPLICANT: RITTER, JOSEPH K.

TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION

TITLE OF INVENTION: THEREIN.

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSER: CUSHMAN DARBY & CUSHMAN

STREET: 1615 L STREET, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20036-5601

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/00282

FILING DATE: 19920110

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SCOTT, WATSON T.

REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 531 amino acids

TYPE: AMINO ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US92-00282-6

Query Match 6.2%; Score 138; DB 5; Length 531;

Best Local Similarity 19.2%; Pred. No. 1.4e-05;

Matches 91; Conservative 72; Mismatches 169; Indels 142; Gaps 20;

2Y 6 VIAPPPSHVRLQNAELVARGHVTTFQOHDCKALVTGSDIGFQTVGLQTHPPGSL 65
DB 29 LVFPWESHWLSMDVRELHARGH-----QAVVLAPEV-----TVYMKGEFP 71
2Y 66 HLLHLAHP-----LG-----PSMLRLINEMARTSDMLCRELPAAFH- 102
DB 72 FTLQTYAFPTYKEEYQREILGNACKGPEPQHEVKTFFETWASIKKFFDLVANSACALLHN 131
2Y 103 -----ALQIEGVVDQMEPAGAVAEASGLPFV-----SVACAL-----PLNRE 141
DB 132 KTLIQQLNSSFDVLTDPVFPFCALLAKVLOIPAVFFLSVFCGDIYATQCPKSSVI 191
2Y 142 PGLPLAVMPEFYGTSDAARERTYTSSEKIYDMLARRHDRVIAHACRGLAPREKILHCFPS 201
DB 192 ENL-LTML-----SDHMTFLQRVKMLYPLTLK-----YICHSITPYESL---AS 233
2Y 202 PLAQISQILPEL-----DFFRKALPDCFHVG-----PLRQ-----232
DB 234 ELLQREMSLVLELHSHASWLVPRGDFVDFYPRIMPNNVFGGNCVKKPLSQEFPEAYN 293
2Y 233 ---PQG---TPGSSSTSYFPSPDKPRIFASLGTQGHRYGLFRTIAKACEEVDALIASRTPLAL 285
DB 294 ASGEGHGIWPSLGSWVSEIPEKKAMEIAEALG-----RIPQTLWRY 335
2Y 286 CGGISATQAGELARGSDIOVDFPADQSAALS--QAQLTITHGGMNTVLDALIASRTPLAL 343
DB 336 TG-----TPSPSLAK--NTILVWLPQNDLLGHKPARAPITHSGSHGIYSGICNGVPMVM 369
2Y 344 PLAPDQPCQVASRIVVHIGIKRASRF--TTSALARQIRSLTNTDYPQRMTKIOA 396
DB 390 PLFGQDQMDNAKMETRGAGVTLNVLEMTADDDLEALNTVNNKSYKENIMELSS 443

RESULT 14
US-08-926-258-2
; Sequence 2, Application US/08926258
; Patent No. 5871983
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Solenberg, Patricia J.
; APPLICANT: Treadway, Patti J.
; TITLE OF INVENTION: Glucosyltransferase Gene gtfE From
; TITLE OF INVENTION: Amycolatopsis orientalis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,258
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: P-10430
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-926-258-2
Query Match 6.2%; Score 137.5; DB 2; Length 408;
Best Local Similarity 24.3%; Pred No. 1e-05;
Matches 109; Conservative 43; Mismatches 201; Indels 95; Gaps 24;
QY 15 VRALQNLAQELVARGHVTTFQOHDCKALVTGSDIGFQTVGLQTHPPGSLHLLHLAHP 74
DB 14 VERLVALAVLRERGAERVMCAPDPCADRLAEVDVPHPLGASARPSAG-----QAKP 66
QY 75 L-GPSMLRLINEMARTSDMLCRELPAAFH---ALQIEGVVDQMEPAGAV---VAEASG 126
DB 67 LTADMLRFTTETIATQ---FERIPAAAECAAVTTTGLL-----AAAGVRSVAEKL 117
QY 127 LP-FVSACALPLNR---EPGLPLAVMPEFYGTS-DAARERTYTSSEKIYDMLARRHDRV 181
DB 118 IPYFVGTHCPYSVPYAPPPPLGEPAPDGDIDQALWERNQS-----AYRYCEPL 171
QY 182 AHHACRGLAPREKILH-HCPSPLAQISQILPELDPRKALPDCFHAVGLPRLQPGTSGSS 240
DB 172 NSRRAAIGLPPVEDIFGHGYT-----DHPWMAADP---VLAPL-QPTDLDVAVQ 215
QY 241 TSYFPSPDKPRIFASLGTQGH-----RYGLFRTIAKACEEVDALIASRTPLAL 293
DB 216 TGAMILPDERPISAELEAFLDAGAPPVYVGLGSLRAPADAA-KVAIEAIRAH-----266
QY 294 AGE---LARG-GDIQVVD-----PA-----DQSAALSQAQLTITHGGMNTVLDALIASRTPL 340
DB 267 -GHRVILSRGWADLVLPDDREDCFAGEVNCQVLFRRVAIVHGGAGTTHVATRAGVPQ 325
QY 341 LALPLATDQPCQVASRIVVHIGIKR---ASRFTTSHALARQIRSLTNTDYPQRMTKIOA 396
DB 326 ILVQLADQPYVYARVAELGVGVHAGDGTPTTDT-----LSAALTALAPETRVRAEA 378
QY 397 ALRL--AGGTPAAADIVEQAMRTCPVL 422
DB 379 VAEVTQTDGAADLLFAAATGNQPAV 406

RESULT 15
US-08-926-053-2
; Sequence 2, Application US/09120053
; Patent No. 5932464
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Solenberg, Patricia J.
; APPLICANT: Treadway, Patti J.
; TITLE OF INVENTION: Glucosyltransferase Gene gtfE From
; TITLE OF INVENTION: Amycolatopsis orientalis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,053
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,258
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872

REFERENCE/DOCKET NUMBER: P-10430

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3334

TELEFAX: 317-276-3861

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 408 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

-09-120-053-2

Query Match 6.2%; Score 137.5; DB 2; Length 408;
 Best Local Similarity 24.3%; Pred. No. le-05;
 Matches 109; Conservative 43; Mismatches 201; Indels 95; Gaps 24;

15 VRALQNLAQELVARGHRVTFQOHDCKALVTGSDIGFTVGLQTHPPGSLSHLHLAAHP 74
 14 VEPLVALAVLRERGAERVCMCAPPCADRLAEVDVPHPLGASARPSAG-----QAKP 66

75 L-GPSMLRLINEMARTSMLCRLPAAPH---ALQIEGVIVDQMEPAGV-----VAEASG 126
 67 LTAEDMLRFTTETIATQ---FERIPAAAECAAVVTTGL-----AAAGVRSVAEKLG 117

127 LP-FVSVACALPLNR---EPGLPLAVMPFEYGTG-DAARERYTTSEKIYDMLMRHDEVI 181
 118 IPYFYGFHCPSYVPSYPYAPPPIGSPAPDGTDIQALWERNQS-----AYRRYGEPL 171

182 AHACRMGLAPREKLH-KCFSPQLAQISQLIPELPPRKALPDCPHAVGPLRQPGQTFGSS 240
 172 NSRRAAIGLPPVEDIFGRGT-----DHPWMAADP---VLAPL-QPTDLDAVQ 215

241 TSYFSPDKPRIFASLGLTQGH-----RYGLERTIAKACEEVDQALLAHCGGLSATQ 293
 216 TGAWILDPERPISAELEAFLDAGAPPVYLGGSLRAPADAA-KVAIEAIRAH----- 266

294 AGE---LARG-GDIOVD-----FA---DQSAALSQAQLTIHGMNTVLDIASSTPL 340
 267 -GHERVILSGWADLVLPDDREDCEFAIGEVNQVLFRVAAVHHCAGAGTTHTVATRAGVPQ 325

341 LALPLAFDPQGVASRIYVHGIGKR-----ASRFTTSHALAQIRSLTNTDYPQMTKIQ 396
 326 ILVPOIADQPYAARVAELGVGVADHGPTPTFDT-----LSAALTNALAPETRVRAEA 378

397 ALRL--AGGTPAAADIVEQAMETCQPV 422
 379 VAETVQTDGAADLLFAAVTGNQPAV 406

Search completed: February 29, 2004, 14:54:58
 > time : 16.0973 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:26:38 ; Search time 50.9584 Seconds
(without alignments)

2389,754 Million cell updates/sec

file: US-09-941-947a-28

irfect score: 2231

quence: 1 MSHFAVIAPFFSHVRLQN.....EQAMTCQPVLSGQDYATAL 431

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

tal number of hits satisfying chosen parameters: 1586107

ninum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2231	100.0	431	5	AAE22312 Pantoea s
2	2231	100.0	431	6	AAO16018 Pantoea s
3	2231	100.0	431	6	ABP96686 Pantoea s
4	1790	80.2	431	2	AAO7464 Polypepti
5	1790	80.2	431	2	AAW87890 Protein e
6	945	42.4	399	2	AAI13989 Zeaxanthi
7	247	11.1	412	4	AAE61974 S. avermi
8	201	9.0	397	6	ABU11976 M. echino
9	183	8.2	432	6	ABM63288 Photorhab
10	183	8.2	470	6	ABP57697 Saccharop
11	181.5	8.1	540	4	ABE64912 Drosophil
12	178.5	8.0	492	4	ABE84216 Amino aci
13	174.5	7.8	541	6	ABU89759 Protein d
14	174.5	7.8	541	7	ADD48913 Human pro
15	171.5	7.7	414	2	AAE93982 Saccharot
16	170	7.6	527	6	ABU54566 Human NOV
17	167	7.5	422	3	AAV83786 S. fradia
18	167	7.5	527	4	AAU29284 Human PRO
19	167	7.5	527	5	AAU77927 Human dru
20	167	7.5	527	5	ABP71237 Human 326
21	167	7.5	527	5	AAE15434 Human 326
22	167	7.5	527	6	ABU58660 Human PRO
23	167	7.5	527	6	ABU88208 Novel hum
24	167	7.5	527	6	ABU84523 Human sec
25	167	7.5	527	6	ABR66397 Human sec

26	167	7.5	527	6	ABR65787 Human sec
27	167	7.5	527	6	ABU99727 Human sec
28	167	7.5	527	6	ABU82966 Human PRO
29	167	7.5	527	6	ABU90087 Novel hum
30	167	7.5	527	6	ABR68136 Human sec
31	167	7.5	527	6	ABU96389 Novel hum
32	167	7.5	527	6	ABU92820 Human sec
33	167	7.5	527	6	ABO08897 Human sec
34	167	7.5	527	6	ABO02949 Human sec
35	167	7.5	527	6	ABR75103 Human sec
36	167	7.5	527	6	ABR94865 Human sec
37	167	7.5	527	6	ABU85838 Human PRO
38	167	7.5	527	6	ABU98998 Novel hum
39	167	7.5	527	6	ABU98213 Novel hum
40	167	7.5	527	6	ABU91919 Novel hum
41	167	7.5	527	6	ABU89612 Human PRO
42	167	7.5	527	6	ABU86453 Human sec
43	167	7.5	527	6	ABU67666 Human sec
44	167	7.5	527	6	ABU80694 Human PRO
45	167	7.5	527	6	ABR99612 Human sec

ALIGNMENTS

RESULT 1

AAE22312
ID AAE22312 standard; protein; 431 AA.

XX AAE22312;

DT 25-JUL-2002 (first entry)

DE Pantoea stewartii zeaxanthin glucosyl transferase (CrtX) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; zeaxanthin glucosyl transferase; CrtX.

XX Pantoea stewartii.

XX WO200218617-A2.

XX 07-MAR-2002.

PP 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-0229858P.

PR 01-SEP-2000; 2000US-0229907P.

PR (DUPO) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

PI Odom JM, Picataggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

DR N-PSDB; AAD35510.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates.

XX Claim 25; Page 136-137; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds. The method comprises a transformed mycoplasma host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the control of regulatory sequences, and contacting the host cell with carbon substrate to produce a carotenoid compound. The method is useful for producing carotenoid compounds such as antheraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in

C the carotenoid biosynthetic pathway and which metabolise single carbon
C substrates. The carotenoids have potent anti-oxidant properties useful in
C diet, and aquaculture elements. The carotenoids are also useful as
C intermediates in the synthesis of steroids, flavours and fragrances and
C compounds for potential electro-optic applications. The present sequence
C is Pantoea stewartii zeaxanthin glucosyl transferase (CrtX) enzyme used
C in the invention

X
Q Sequence 431 AA;

Query Match 100.0%; Score 2231; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.4e-215;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 MSHFAVIAPPPFVSHVRLALQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
b 1 MSHFAVIAPPPFVSHVRLALQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
Y 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120
b 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120
Y 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFEYGTSDAARERTTSEKIYDLMRRHDRV 180
b 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFEYGTSDAARERTTSEKIYDLMRRHDRV 180
Y 181 IAHAACRMGLAPREKLUHCFSPALQISQLIPELDPFRKALPDCFHAGVPLRPOQGTGSS 240
b 181 IAHAACRMGLAPREKLUHCFSPALQISQLIPELDPFRKALPDCFHAGVPLRPOQGTGSS 240
Y 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
b 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
Y 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360
b 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360
Y 361 IGKASFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMTCOP 420
b 361 IGKASFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMTCOP 420
Y 421 VLSQDYATATL 431
b 421 VLSQDYATATL 431

RESULT 2

AA016018 standard; protein; 431 AA.

AA016018;

20-FEB-2003 (first entry)

Pantoea stewartii zeaxanthin pyrophosphate synthase.

Carotenoid; crt.

Pantoea stewartii.

WO200279395-A2.

10-OCT-2002.

25-JAN-2002; 2002WO-US002124.

26-JAN-2001; 2001US-0264329P.

04-MAY-2001; 2001US-028994P.

(CRGI) CARGILL INC.

De Souza ML, Kollmann SR, May CA, Schroeder WA;

XX WPI; 2003-075455/07.
DR N-PSDB; ABT14190.

XX Novel isolated nucleic acid useful e.g. to engineer host cells with the
PT ability to produce particular carotenoids and polypeptides useful in cell
PT -free systems to make particular carotenoids.

XX Claim 7; Page 58-59; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of
CC carotenoid (crt)-related proteins. The crt-related DNA and protein
CC sequences of the invention are useful for engineering cells which are
CC able to produce carotenoids. The present amino acid sequence represents a
CC crt-related protein of the invention
XX
SQ Sequence 431 AA;

Query Match 100.0%; Score 2231; DB 6; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.4e-215;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSHFAVIAPPPFVSHVRLALQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
DB 1 MSHFAVIAPPPFVSHVRLALQALVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
QY 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120
DB 61 PGSLSHLHLAAHPLGSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMPAGAV 120
QY 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFEYGTSDAARERTTSEKIYDLMRRHDRV 180
DB 121 VAASGLPFFVSVACALPLNREPGPLVAVMPFEYGTSDAARERTTSEKIYDLMRRHDRV 180
QY 181 IAHAACRMGLAPREKLUHCFSPALQISQLIPELDPFRKALPDCFHAGVPLRPOQGTGSS 240
DB 181 IAHAACRMGLAPREKLUHCFSPALQISQLIPELDPFRKALPDCFHAGVPLRPOQGTGSS 240
QY 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
DB 241 TSVPSPDKPRIFASLTGLQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARG 300
QY 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360
DB 301 GDIQVDFADQSAALSOAQLTIHGGMNVTLDIAISRTPLALPLAFDQPGVASRIVYHG 360
QY 361 IGKASFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMTCOP 420
DB 361 IGKASFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGTTPAAADIVEQAMTCOP 420
QY 421 VLSQDYATATL 431
DB 421 VLSQDYATATL 431

RESULT 3

ABP96686 standard; protein; 431 AA.

ABP96686;

03-JUN-2003 (first entry)

Pantoea stewartii zeaxanthin glucosyl transferase SEQ ID NO:4.

Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
KW crtI; crtB; crtZ; zeaxanthin glucosyl transferase; enzyme; phytoene;
KW carotenoid.

Pantoea stewartii.

WO2003016503-A2.

XX

27-FEB-2003.
 15-AUG-2002; 2002WO-US026647.
 15-AUG-2001; 2001US-0312646P.
 (DUFO) DU PONT DE NEMOURS & CO E I.
 Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PB,
 WPI; 2003-269323/26.
 N-PSDB; ACC44760.
 Novel nucleic acid molecule isolated from *Pantoea stewartii* encoding a carotenoid biosynthetic enzyme, useful for regulating carotenoid biosynthesis in an organism.
 Claim 4; Page 59-60; 58pp; English.
 The present invention describes *Pantoea stewartii* carotenoid biosynthetic enzymes (I). More specifically described are the geranylgeranyl pyrophosphate synthase (crtB), zeaxanthin glucosyl transferase (crtX), lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase (crtE) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating carotenoid biosynthesis in an organism, by over-expressing (I) in an organism, such that the carotenoid biosynthesis is altered in the organism. (II) and the genes encoding (II) are useful for converting phytoene to the carotenoids, for creating recombinant organisms that have the ability to produce various carotenoid compounds, and also for enhancing or manipulating carotenoid compounds. (I) can also be used for producing gene products having enhanced or altered activity

Sequence 431 AA;

Query Match 100.0%; Score 2231; DB 6; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.4e-215;
 Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MSHFAVIAPPPSHVRALONLAQELVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
 1 MSHFAVIAPPPSHVRALONLAQELVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 121 VAEASGLPVSVCALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 121 VAEASGLPVSVCALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 181 IAHACRMGLAPREKHLHCFSPLAQISQILPELDFPRKALPDCFHAVGLPQPGTSS 240
 181 IAHACRMGLAPREKHLHCFSPLAQISQILPELDFPRKALPDCFHAVGLPQPGTSS 240
 241 TSFFSPDKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLAHCGGLSATORAGELARG 300
 241 TSFFSPDKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLAHCGGLSATORAGELARG 300
 301 GDIQVVDFAQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 301 GDIQVVDFAQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 361 IGRASRFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGGTAAADIIVEQAMRTCP 420
 361 IGRASRFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGGTAAADIIVEQAMRTCP 420
 421 VLSGQDYATAL 431
 421 VLSGQDYATAL 431

RESULT 4

AAR07464
 ID AAR07464 standard; protein; 431 AA.
 AC XX
 AAR07464;
 AC XX
 DT 24-OCT-2003 (revised)
 DT 28-JAN-1991 (first entry)
 XX
 DE Polypeptide with enzymatic activity for the conversion of zeaxanthin into zeaxanthin diglucoside.
 DE
 XX Carotenoid biosynthesis; vitamin A; cancer; food coloring.
 XX
 OS *Pantoea ananatis*.
 XX
 PN EP9393690-A.
 XX
 PD 24-OCT-1990.
 XX
 PF 20-APR-1990; 90EP-00107493.
 XX
 PR 21-APR-1989; 89JP-00103078.
 PR
 XX 05-MAR-1990; 90JP-00053255.
 XX
 PA (KIRI) KERIN BEER KK.
 XX
 XX Misawa N, Kobayashi K, Nakamura K;
 PI WPI; 1990-322212/43.
 XX
 DR N-PSDB; AAQ06294.
 DR
 XX DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.
 PT
 PT Claim 2; Fig 2; 40pp; English.
 PS
 XX Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)
 CC
 CC
 XX Sequence 431 AA;
 SQ
 Query Match 80.2%; Score 1790; DB 2; Length 431;
 Best Local Similarity 80.5%; Pred. No. 9.2e-171;
 Matches 347; Conservative 29; Mismatches 55; Indels 0; Gaps 0;
 QY 1 MSHFAVIAPPPSHVRALONLAQELVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
 DB 1 MSHFAVIAPPPSHVRALONLAQELVARGHRVTFQOHDCKALVTGSDIGFOTVGLQTHP 60
 QY 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 DB 61 PGSLSHLHLAAHPGLSPMLRLINEMARTSDMLCRELPAAFHAIQIEGVIVDQMEPAGAV 120
 QY 121 VAEASGLPVSVCALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 DB 121 VAEASGLPVSVCALPLNREPGLPLAVMPFYGTSDAARERTTSEKIYDWLMRRHDRV 180
 QY 181 IAHACRMGLAPREKHLHCFSPLAQISQILPELDFPRKALPDCFHAVGLPQPGTSS 240
 DB 181 IAHACRMGLAPREKHLHCFSPLAQISQILPELDFPRKALPDCFHAVGLPQPGTSS 240
 QY 241 TSFFSPDKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLAHCGGLSATORAGELARG 300
 DB 241 TSFFSPDKPRIFASLGTQGHRYGLFRTIAKACBEVDQAQLLAHCGGLSATORAGELARG 300
 QY 301 GDIQVVDFAQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 DB 301 GDIQVVDFAQSAALSOAQLTTHGGMNTVLDIAISRTPLALPLAFDPQGVASRIYVHG 360
 QY 361 IGRASRFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGGTAAADIIVEQAMRTCP 420
 DB 361 IGRASRFTTSHALAQIRSLTNTDYPQMTKIQAALRLAGGTAAADIIVEQAMRTCP 420

X 10-OCT-2002.
X 28-NOV-2001; 2001WO-US044285.
X 28-NOV-2000; 2000US-00724797.
X (SLOK) SLOAN KETTERING INST CANCER RES.
X Thorson J;
X WPI; 2003-052897/08.
X N-PSDB; ABX56038.
X Novel nucleic acid molecule from nonchromoprotein enediyne biosynthetic
X gene cluster from Micromonospora echinospora useful for conferring
X calicheamicin resistance on a subject.
X Claim 25; Page 100-101; 179pp; English.
X The present invention relates to the isolation of the Micromonospora
X echinospora spp. calicheamicin biosynthetic gene cluster
X encoding proteins and enzymes used in the biosynthetic production of
X calicheamicin, including aryltetraaccharide and aglycone. The gene
X cluster also includes the gene encoding for the protein that confers
X calicheamicin resistance. The calicheamicin biosynthetic gene cluster is
X a nonchromoprotein enediyne biosynthetic gene cluster. Expression vectors
X containing genes from the biosynthetic gene cluster are also disclosed.
X The expression vectors are useful for producing calicheamicin
X biosynthetic proteins. The calicheamicin self-resistance gene provides an
X approach for gene therapy, for example, by introduction of enediyne
X resistance genes into bone marrow cells, thus increasing resistance and
X allowing tolerance to chemotherapeutic doses of calicheamicin. ABU11964-
X ABU12010 represent proteins and enzymes encoded by genes in the M.
X echinospora calicheamicin biosynthesis gene cluster. (Updated on 23-OCT-
X 2003 to standardise OS field)
X Sequence 397 AA;
Query Match 9.0%; Score 201; DB 6; Length 397;
Best Local Similarity 21.2%; Pred. No. 4.1e-11;
Matches 96; Conservative 69; Mismatches 184; Indels 104; Gaps 14;
2Y 1 MSHEAVIAPPPFSSVRALQNLQAEVARGHRTVP-----FQGH--- 38
2b 1 MAHLIVNVASHGLILTLVTVELRGHVSIVTAGFAEPVRAAGATVPVQSEIID 60
2Y 39 -DCKALVTGSDIGFTVGLQTHPPGSLHLHLAAHPLGFSMLRLINEMARTSDMLCREL 97
2b 61 ADAAEVFGSDDLGVR-----PHLMYLRN-----VSVLRATAEALGDGV 99
2Y 98 PAAFHAIQIEGVVDQXE-PAGAVVAESGLPFVSVACALPLNRPGI----- 144
2b 100 P-----DLVLVDPPFIAGQLAARWRPAPVLSAAFSNHEVSFSDMYTLAGTID 151
2Y 145 PLAVMPFEYGTSDAARERTYTSKIDYLMARRHDRVIAHACRGLAPRLKHLHCFSP 204
2b 152 PLDELVFRDTRLDLAELHGLSRVVDWC-----NHV-----BQLNLVFPVKA 193
2Y 205 QISQILPDLFPRKALPCFHANGVLQPCQTPGSSITSYSPSPDKPIFASLGLQCHRY 264
2b 194 -----FQIAGDTDFDRFVFGVGFCDRDFLGEWTR--PADLPLVVLVSLGTTNDRP 243
2Y 265 GLFRTIAKACEVDAQLLHACGGLSATQAGELAGGDIQVDFADQSAALSQAQLTITH 324
2b 244 GFFRCARAFDQGPWEVVMVTLGGQVDPALGDLP--PNVEAHRVWPVHVKLEQATVCVTH 301
2Y 325 GGMVTVDAIASRTPLLPALPAFADPGVASRIVVHGI-----GKRSEFTTSHALAQIR 379
2b 302 GGMGLTMEALYNGRPELVVVPQSESDVQPNRRVQDGLGAVLPGEKAGDT-----LLAANG 357
2Y 380 SLITNTDVPQMTKIOAALRLAGGTTPAAADIVE 412
2b 380 SLITNTDVPQMTKIOAALRLAGGTTPAAADIVE 412

Db 358 AVAADPALLARVEAMRGHVVRAGGARAADAVE 390
RESULT 9
ABM69288
ID ABM69288 standard; protein; 432 AA.
XX AC
XX ABM69288;
XX DT 20-NOV-2003 (first entry)
XX Photorhabdus luminescens protein sequence #2385.
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
XX antibiotic; biopesticide; virulence factor; disease model; plague;
XX whooping cough.
XX Photorhabdus luminescens.
XX WO20020294867-A2.
XX 28-NOV-2002.
XX 07-FEB-2002; 2002WO-IB003040.
XX 07-FEB-2001; 2001FR-00001659.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX Buchrieser C;
XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX Claim 2; SEQ ID NO 2385; 1205pp; French.
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens proteins
XX Sequence 432 AA;
Query Match 8.2%; Score 183; DB 6; Length 432;
Best Local Similarity 21.8%; Pred. No. 3e-09;
Matches 99; Conservative 62; Mismatches 175; Indels 118; Gaps 19;
QY 5 AVIAPPPFSSHVRLQNLQAEVARGHRTVP-----QQHDKALVTGSDIGFTV 55
Db 7 AAVATP--GHVTPMLKVTSLSITQGHQVTVFTGALFRQVEALNAQPIPFDEQIDFDRH 64

56 LQTH-----PPGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRLPAAFHALQIEG 108
 65 LKQRPDRALPPGHLQALAF-----KDFCSPIP--FLDRQLQ 103
 109 VIVDQEPAGAVVAASGLPVFVSACALPL-----NREPLVAMPPEYGTSD----- 157
 104 IIEHQO--AELLIANC---FYGI---LPLQKKAHSLPVIIGITPLAYSSKDSIFWG 155
 158 -----AARERTYTSKIIDWLMRR-----HDRVIAHACRMGLAPREKLHHCFSF-- 202
 156 PRIPALLPSLTHSQLDEETRLQITVQDSFNDALAQSGC-----TTLTRFNDEV 208
 203 -----LQISOLIFELDFPRKALPDCFAVGPLRQPGTSGSSTSPSPKPRIFASL 256
 209 ILGSDRFLQLSLT--AFYPREELPDTVHFTGPLPNPAPITESPOLWEEDPRFLVITQ 266
 257 GTLQCHR-----YGLPRTIAKACEVDQAQLLHACGGLSATQAGELARGGDIQVVDADQ 311
 267 GTMSNTDLNQLIPFTLRALAKLPVZV-----LATGGSSI-----EQENIPDN 311
 312 SAA-----LSQAQLTIHGGMNTVLDIASTRTPLLALPLAFDQPGVASRIVYHGIG 362
 312 ARIEBFIFGEHMLPKASLIISGGYGTINALNHGTPVLVIAITGEGKOETAFRVWAGCG 371
 363 -KRASFTTSHALAQISLNTDYDQRMKIQ 395
 372 INLDTAQPTESOLKOTVENMLTNLFKORAIQV 405

RESULT 10

3P57697
 ABP57697 standard; protein; 470 AA.

ABP57697;
 22-JAN-2003 (first entry)

Saccharopolyspora ORF LI butenyl-spinosyn biosynthetic gene product.

Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide; metabolite; spinosyn.

Saccharopolyspora sp.

WO200279477-A2.

10-OCT-2002.

28-MAR-2002; 2002WO-US009968.

30-MAR-2001; 2001US-0280175P.

(DOWC) DOW AGROSCIENCES LLC.

Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C, Mitchell JC;

WPI; 2003-058434/05.

N-PSDB; ABV75558.

New butenyl-spinosyn biosynthetic genes, useful for increasing the production of butenyl-spinosyn insecticidal macrolides, or for changing the metabolites or products produced by spinosyn-producing microorganisms.

Claim 1; Page 204-206; 218pp; English.

The invention relates to a novel DNA molecule comprising a DNA sequence that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl-spinosyn biosynthetic genes are useful for increasing the production of butenyl-spinosyn insecticidal macrolides. The genes are also useful for changing the metabolites or products produced by spinosyn-producing

CC microorganisms. The present sequence represents one of the butenyl-spinosyn biosynthetic polypeptides of the invention
 XX
 SQ Sequence 470 AA;

Query Match 8.2%; Score 183; DB 6; Length 470;
 Best Local Similarity 21.0%; Pred. No. 3.3e-09;
 Matches 97; Conservative 80; Mismatches 212; Indels 72; Gaps 15;

QY 1 MGHFAVIAPPPFVSHVRLQNLQAELVARGHRVTPFQHQDCKALV--TGSIDIGFOTVG--- 55
 DB 46 MAHFVPATYADVAHAIIGPLVPSRALVERDQVTTWYGENYRAAVERSGADFAAPVGEHFI 105
 QY 56 -----LQTHPPGSLSHLLHAAHPLGPSMLRLINEMARTSDMLCRLPAAFHALQIEGVIV 111
 DB 106 DGRELEKQFPESIQMSARRARWL-----MDNEWVPAYEGQYRDVAVVDRTRADVILLA 159
 QY 112 DQWEPAGAVVAASGLPVFVSACALPLNREPLVAMPPEYGTSDAARERTYTSSEKIYD 171
 DB 160 DASWGPALKVHAVTGVLTWATISQMPILLPDPVAPPDGTGWKEGTSPPHLLNRNIGRLIN 219
 QY 172 WLMRRHDRVIAHACRMGLAPREKLHHCFT-----SPLAQISOLIFELDPFR 217
 DB 220 ALV--HD-----PCMKKINAFMNSIGVPSREVSESPYLFMQAGTRSLSEYPR 264
 QY 218 KALPDCFAVGPLRQPGTSGSSTSYFBS-----PDKPRIFASLGTQGHRYGLFRITIA 271
 DB 265 -ALPQOMHFIGRL-EFDSPMGVG---LFSWNGELDGRPVVLVITQGTWAVDADDLIRPAL 319
 QY 272 XACEEYDQAQLLHACGGLSATQAGELARGGDIQVVDADQSAALSQAQLTIHGGMNTVL 331
 DB 320 RGL--AGDQVLVVATTGREGVDLGVP--DNARVASFLPYRELMEKLAADVTTNGGFGTVQ 375
 QY 332 DAISRTPELLALPLAFDQPGVASRIVYHGIG-KRASRFTTSHALAQISLNTDYDQOR 390
 DB 376 QALSHGLPLVWAGRSSEKTDVCAVWAGSGVDLSTRPSPQQVAGAVKVMSTDPYR-- 433
 QY 391 MTKIOAALRLA-----GCTPAADIVDAQMETCOPVLSGQ 425
 DB 434 ----QAAQLAVEAYEYDACGTAVKLLRELATRRPVIASR 470

RESULT 11

AB864912
 ID AB864912 standard; protein; 540 AA.

AC AB864912;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 21528.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKS) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09015.

NT New isolated nucleic acid detection reagent for detecting 1000 or more
 TT genes from Drosophila and for elucidating cell signaling and cell-cell
 TX interactions.

XX Disclosure: SEQ ID NO 21528; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signaling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 XX sequences (ABU161840-ABU16175) and the encoded proteins (ABU57737-
 XX ABU572072). The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 540 AA;

Query Match 8.1%; Score 181.5; DB 4; Length 540;
 Best Local Similarity 23.7%; Pred. No. 5.8e-09;
 Matches 109; Conservative 66; Mismatches 202; Indels 83; Gaps 19;

Y 7 IAPFFSHVRAALQNAELVARGHRTFFQHQCKALVTG-----SD 48
 b 36 IPSPF-----QKVRPLIALVERGHKVTWTPADYPAKIDGVRHVRPMLNQLMKNMD 91
 Y 49 IGFVTGLQHPGSLHLLHAHPGLPSMLRLINEMARTSMLCRELPAAFPALQIEG 108
 b 92 QFFDALGDKWRGVLTIFVNSHAI-----LNNNGVQ---MLMRDKSIRPDMIMVEA 142
 Y 109 VIVDQMEPAGAVVAAAGLFPVSACA-----LPLNREPCGLPVMFEFGTSDAAR 160
 b 143 SHLDALYG-----LAEFYNTALLGISEMWTWHIDYLAGNLAPSIVYPIGPNFALDNTFL 198
 Y 161 ER-----YTTSEKIYDMLRHRHVRVIAHACRMGLAPREKLEHCFSPLAQISOLIPELDP 215
 b 199 SRWNWIIYIEKLERLVRPAQVRLFK--KFGYPAEKIDEL---RAFSVILVNSHF 253
 Y 216 P-----RKALPDCFHAVG--PIRQPGTGGSS--TSYFSPDKPRIPASLGLTQGHRYGLPT 269
 b 254 SMGRVRANVPNIIEVAGVHLSEPPPCGABLOKYLDEAERGAIFYSMGQDILIKY----- 308
 Y 270 IAKACEEVDQAOLLAHCGLSATO-----AGELA-----RGDIOQVDFPADQSAALS--QAQ 319
 b 309 ---LPENKQKOLLVF---LQMKORVTKSELSMLANKSENIYVDKVPQWVLAHPNLR 362
 Y 320 LTITHGGMNTVDAASRTPLALPLAPDQGVASRIVYHIGKGRAS--RTTSHALARO 378
 b 363 LFITHGGLQSWEAIDNGVPMGLPFLFDQFNNIHRYQLAGMAKVLDPNLDNADTLIETI 422
 Y 379 RSLLTNTDYPQRTKIOALRLAGGTAAADI--VEQAMR 416
 b 423 KELLNPSYAORAKEMAAFPDRPMSPLDTAIWTTETALR 462

RESULT 12

AB84216

ID AB84216 standard; protein; 492 AA.

XX

AC AB84216;

XX

DT 06-AUG-2001 (first entry)

XX

DE Amino acid sequence of an UDP-glucose:aglycon-glucosyltransferase.

XX

KW UDP-glucose:aglycon-glucosyltransferase; UDP-GAG; cyanohydrin; terpenoid;

KW glucose; transgenic plant; cyanogenic glucoside biosynthesis;

XX pathogen resistance; herbivore response.

OS Sorghum bicolor.

XX

PN WO200140491-A2.

XX 07-JUN-2001.

PD XX

XX 29-NOV-2000; 2000WO-EP011982.

XX 01-DEC-1999; 99EP-00123838.

XX (LUMI-) LUMINIS PTY LTD.

XX (UYRO-) UNIV ROYAL VETERINARY & AGRIC.

XX Hoej P, Moeller BL, Jones PR;

XX WPI; 2001-374846/39.

XX N-PSDB; AAF90242.

XX DNA molecule coding for UDP-glucose:aglycon-glucosyltransferase

XX conjugating cyanohydrin, terpenoid or phenyl derivative to glucose, for

XX producing transgenic plants having modified cyanogenic glucoside

XX biosynthesis.

XX Claim 7; Page 27-29; 31pp; English.

XX The present sequence represents a UDP-glucose:aglycon-glucosyltransferase

XX (UDP-GAG) polypeptide. The enzyme conjugates a cyanohydrin, terpenoid,

XX phenyl derivative or hexanolderivative to glucose. UDP-GAG polynucleotides

XX are useful for producing transgenic plants having modified cyanogenic

XX glucoside biosynthesis. Constitutive, inducible or tissue-specific

XX expression of UDP-GAG is useful for obtaining transgenic cyanogenic

XX plants with altered resistance to pathogens and herbivore responses

XX Sequence 492 AA;

XX Query Match 8.0%; Score 178.5; DB 4; Length 492;

XX Best Local Similarity 23.2%; Pred. No. 1e-08;

XX Matches 113; Conservative 73; Mismatches 191; Indels 111; Gaps 22;

QY 3 HFAVIAPFFSHVRAALQNAELVARGHRTFFQHQCKALVTGSDIGPQTGLQTHPP 61

Db 12 HWLVPPFGQGVNAPLQMLARLLHARGARVTFYTYQTYNRLLRAK--GEANV---RPP 65

QY 62 GSLSHLLHAAHPGLPSMLRLINEMARTSMLCRELPAAFPAL-----QIEG----- 108

Db 66 ATSSARFRIEVIDDGLSLVQNDVGLVDSLRKNCCLHPFRALLRLRGQVEGODAPPVT 125

QY 109 -VIVDQMEPAGAVVAAAGLFP-----FVSVACAL--PLNREPGLPVMFEFGTSDAARE 161

Db 126 CWGDDVWMTFAAAAAREAGIPEVQFTASACGLGLYHLYGELVERGLVPER-DASLLADD 184

QY 162 RYTTSEKIYDML-----MRHRDRIAHHACK-----MGLAPREKLEHCFSPLAQISOL 209

Db 185 DYLDTP--LEWVPGKSHMLRLEDMPT---FCRTTDPDDVWVSATLQOMESAAGSKALIINT 239

QY 210 IPELDFPRKALPDC-----FHAVGFLRQPGTGGSTSYFSPD----- 248

Db 240 LYLELE---KQVVDALAAFFPPIYTVGPLAEVIASSDSASAGLAAMDISIWQEDTRCLSNL 296

QY 249 --KPR---IPASLGTGLQGHRYGLPRTTAKACEEVDQAOLLAHCG----- 288

Db 297 DGKSAGSVVYVNFSGM-----AVMTAAQAREFALGLASCSPLWVKRPDQVVEGEE 347

QY 289 --LSATQAGELARGSDIQVDFPADQSAALSQAQ--LTITHGGMNTVDAASRTPLIALP 344

Db 348 VLLPEALLDEVARGGL-VVWPCPQAAVLKHAAGVLFVSHCGWNSLLEATAAGQPVILAWP 406

QY 345 LAFQDPGVASRIV-YHIGKGRASRTTSHALAROIRSLLTNTDYPQRTMK-----IOA 396

Db 407 CHGEQNTNCRQLCEVWGNGAQLPREVESGAVARLVREKVMVGDIGKEKRAAEWKAAREA 466

QY 397 ALRLAGGT 404

Db 467 AARKGAS 474

RESULT 13
ABU89759
ABU89759 standard; protein; 541 AA.
ABU89759;
10-JUL-2003 (first entry)
Protein differentially expressed in cardiovascular disease #53.
Cardiovascular disease; arteriosclerosis; ischaemia; angina pectoris;
myocardial infarction; cardiast; antiarteriosclerotic; antianginal;
gene therapy; differential gene expression.
Homo sapiens.
WO2003031650-A2.
17-APR-2003.
02-OCT-2002; 2002WO-EP011034.
08-OCT-2001; 2001GB-00024145.
(FARB) BAYER AG.
Munnes M, Gehrman M, Wick M, Schmitz G;
WPI; 2003-403108/38.
N-PSDB; ACAB9932.
Predicting, diagnosing or prognosing a cardiovascular disease, e.g.
angina, ischemia, myocardial infarction or arteriosclerosis by detection
of a polynucleotide in a biological sample comprises detecting a
hybridization complex.
Claim 3; Page 406-408; 454pp; English.
The invention describes a method of predicting, diagnosing or prognosing
a cardiovascular disease by detection of a polynucleotide in a biological
sample comprises hybridising at least one of the polynucleotide to a
nucleic acid material of a biological sample, thus forming a
hybridisation complex, and detecting the hybridisation complex. The
polynucleotides, polypeptides, antisense molecule, antibody and reagent
are useful for preparing compositions for preventing, predicting or
diagnosing, or a medicament for treating a cardiovascular disease, e.g.
arteriosclerosis, ischaemia, angina pectoris, or myocardial infarction.
This sequence represents a protein identified in the invention a being
differentially expressed in individuals with cardiovascular disease
Sequence 541 AA;
Query Match 7.8%; Score 174.5; DB 6; Length 541;
Best Local Similarity 20.1%; Pred. No. 2.9e-08;
Matches 96; Conservative 71; Mismatches 173; Indels 137; Gaps 19;
6 VIAPP--PFSHVRALONLAQELVARGHRTVFFQHDCKALVTGSDIG--FQTVGLQTHP-- 60
24 IIVPTMFESHMYIFKTLASALHGHGTFL-----LSEGRDIAPSNHYSLQRYPGI 76
61 -----PGSLSHLLHLAAHPLGSPMLRINEMARTSDMLCRELPAAFH 102
77 FNSITSDAFLOSMEINISGRLLTAI-----ELFDILDHYTKCDLM-----VGNH 121
103 ALQIEG-----VIVDMEPAGAVARASGLPFVSVACALPLNRSPG--LPLAVMPFE 152
122 AL-IQLKKEKEDLLVDPNDMCGFVAHLLGVKAYVESTGLWYPAEYVGPAPLAYVP-E 179
153 YGTSDAAREYTTSEKIDVLMNRHDRVIAHACWGLA----PREKLHCFSPLAIOISQ 208
180 FNSLITDMN-----LLQRMKNTGVLIISKLGVFLVLPK-----YERIMQYN 223
209 LIPE-----LDFPRKALPDCFHAVGPLRQPGQPGSGSTSYFFSPD 248

DB 224 LLEFKSMYDLVHGSSLMMLCTDVALEFPFPLPNVYVGGILTKPAS-----PLPE 274
QY 249 KPRIFASLGTQGHRYGLFRTIAKACEVDQAOLLAHCGGLSATQAGELAR----- 299
DB 275 DIQRWVNGANEHGFVLVSFGAGVYLSSEDIANKL-----AGALGRLPQKVIWRF 323
QY 300 -----GGDIOVVDFADQSNALSOAQLT--ITHGGMNTVLDALASRTPLLALPLAFDQ 349
DB 324 SGPKPKNGNNTKLIEWLPQNDLLGHSKIKAFLSHGGLNSIFETWYHGVVPGVPLGDH 383
QY 350 PGVASRIYVHIGIGKAS--RFTTSHALAQIRSLTNTDYPQRMTKIQALRLAGTTP 405
DB 384 YDTMTRVQAKGKMGILLEWKTIVTEKELYEALVKVINPNPSYRQRAQKLSIHKDQPGHP 440

RESULT 14

ADD48913
ID ADD48913 standard; protein; 541 AA.

XX AC ADD48913;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein Q16880, SEQ ID NO 14624.

XX KW Human; pain; neuronal tissue; gene therapy;

XX KM spinal segmental nerve injury; chronic constriction injury; CCI;

XX KN spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN WO2003016475-A2.

XX XX 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEHO) GEN HOSPITAL CORP.

XX PA (FARB) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX XX WPI; 2003-268312/26.

XX DR GENBANK; Q16880.

XX PT New composition comprising two or more isolated polypeptides, useful for
preparing a medicament for treating pain in an animal.

XX PS Claim 1; Page; 1017pp; English.

XX CC The invention discloses a composition comprising two or more isolated rat
or human polynucleotides or a polynucleotide which represents a fragment,
derivative or allelic variation of the nucleic acid sequence. Also
claimed are a vector comprising the novel polynucleotide, a host cell
comprising the vector, a method for identifying a nucleotide sequence
which is differentially regulated in an animal subjected to pain and a
kit to perform the method, an array, a method for identifying an agent
that increases or decreases the expression of the polynucleotide sequence
that is differentially expressed in neuronal tissue of a first animal
subjected to pain, a method for identifying a compound which regulates
the expression of a polynucleotide sequence which is differentially
expressed in an animal subjected to pain, a method for identifying a
compound that regulates the activity of one or more of the
polynucleotides, a method for producing a pharmaceutical composition, a
method for identifying a compound or small molecule that regulates the
activity in an animal of one or more of the polypeptides given in the
specification, a method for identifying a compound useful in treating
pain and a pharmaceutical composition comprising the one or more

polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung)), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 541 AA;

Query Match 7.8%; Score 174.5; DB 7; Length 541;
 Best Local Similarity 20.18; Pred. No. 2.9e-08;
 Matches 96; Conservative 71; Mismatches 173; Indels 137; Gaps 19;
 6 VIAPP--FFSHVRLQALQAEVLVARGHRVTFQCHDCKALVTSGDIG-FQTGVLQTHP-- 60
 24 IIVPPIMFESHMYIFKTLASALHGERGHTVFL-----LSEGRDIAPSNHYSLQRYPGI 76
 61 -----PGSLSHLLHAAHPLGSMRLINEMARTSDMLCRELPAAFH 102
 77 FNSTTSDAFLOSQRNIFSGRLTAI-----ELFDILDHYTKNCULM-----VGNH 121
 103 ALQIEG-----VVDQMEPAGAVARASGLPFFSVACALPLNREPQ--LPLAVMPFE 152
 122 AL-IOGLKKEKFDLLVDPNDCMGFVIAHLILGVKYAVFSTGLWPAEYVGAPAPLAYVP-E 179
 153 YCTSDARERTTSEKIYDMLRRHRDRIAHACRMGLA----PREKLHHCPSPLAQISQ 208
 180 FNSLLTDRN-----LQKQNTGYLISRLGVSLVLPK-----YERINQKN 223
 209 LIPE-----LDPPKALPDCFHAGVPLQPOQCTPGSSTSYFSPD 248
 224 LLPEKSMYDLVHGSSLMWLCIDVALEFPRTLENVYVGGILTKPAS-----PLPE 274
 249 KPRIFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELAR----- 299
 275 DLQRNVNGANEHGFVNSFGAGVKYLSIEDIANKL-----ACALGRLPQKVINRF 323
 300 -----GSDIQVVFADQSAALSQAQLT--ITHGMNTVLDIAASRTPLALPLAFDQ 349
 324 SGPKPKNLGNNTKLEIWPQNDLLGHKIKAFLSHGGLNSIFETWVGVPVVGIPFGDH 383
 350 PGVASRIVVHIGIKRAS-RFTTSHALARQIRSLTNTDYDQRTKIOAALRAGGTP 405
 384 YDTWTRVQAKGMLLEKWKVTEKELYEALVKVINNPSTQRAQKUSEIHKQOPGHP 440

RESULT 15

AAR93982 standard; protein; 414 AA.

AC AAR93982;

DT 16-OCT-2003 (revised)

DT 15-AUG-1996 (first entry)

XX Saccharothrix aerocolonigenes glycosyl transferase.

XX Glycosyl transferase; glycosylating; indolopyrrolocarbazole;
 XX Actinomycetales; cancer treatment; anti-tumour.

OS Lechevalieria aerocolonigenes; (ATCC 39243).

XX WO9534653-A1.

PD 21-DEC-1995.

XX 31-MAY-1995; 95WO-JP001065.

XX 13-JUN-1994; 94JP-00154127.

PA (BANY) BANYU PHARM CO LTD.
 XX Kojiri K, Suzuki H, Kondo H, Suda H;
 XX WPI; 1996-049691/05.
 DR N-PSDB; AAT29923.

XX Gene coding for glycosyl transferase - useful in glycosylation of
 PT indolo:pyrrolo carbazole derivs.

XX Claim 4; Page 4-6; 45pp; Japanese.

XX AAR93982 is a glycosyl transferase enzyme derived from Saccharothrix
 CC aerocolonigenes (ATCC 39243). The glycosyl transferase (GT) may be from
 CC any organism of the order Actinomycetales. GT is used in a method of
 CC glycosylating indolopyrrolocarbazole derivs. Glycosylated
 CC indolopyrrolocarbazole derivs. are useful as anti-tumour agents. GTs can
 CC be made by recombinant means using new GT-producing strains Streptomyces
 CC lividans TK21 (pNGT207) FERM BP-5091 and S. morbariensis BA 13793
 CC (pNGT207) FERM BP-5090. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 414 AA;

Query Match 7.7%; Score 171.5; DB 2; Length 414;
 Best Local Similarity 21.48; Pred. No. 4e-08;
 Matches 96; Conservative 68; Mismatches 156; Indels 129; Gaps 19;
 6 VIAPFFSHVRLQALQAEVLVARGHRVTF----- 34
 7 VATTFGDGHVPMVPAQAEVMSRGHEVRYTGTAKFRSTVERTGARHEPMDAHDFGMPR 66
 35 ---FQCHDCKALVTSGDIGFQTVGLQTHPQCSLSHLHLAAHPLGFSMLRLINEMARTSD 91
 67 EAPFPQHLGTGTGMIAGFRDIFIE-----PAADQMTDLL----- 102
 92 MLCRELPAAPHALQIEGVIVDQMEPAGAVARASGLPFFSVACAL-----PLNREP 142
 103 ALLEDFPA-----DVLVTDETFFGAGFVSERTGIPVAMIATSIYVFSRDRPAPGL-- 153
 143 GLPLAVMPFEVGTSDAARERTTSEKIYDML-----MERHDRVIAHACRMGLAPREKLH 198
 154 GLPFS-----SSRLGLRLNTVLKQITDRVMDLRRHADVVD---RVGL-PRIR-KG 201
 199 CFSPLAQISQL-----IPELDPPKALPDCFHAGVPLRQPGCTGSGSTSYFSPD--KPR 251
 202 AFENIMRTFDLYLGTVFSPEYPRGDMPEVRFVFPVSPAP2DFTPPAWMGELDSGRPV 261
 252 IFASLTQGHRYGLFRTIAKACEVDQAQLLAHCGGLSATQAGELARGG----- 301
 262 VEVTTQ-----TVANDAE-----RLLPALAAEDVLVATTGAPLEPMPA 306
 302 DIQVVFADQSAALSQAQLTITHGMNTVLDIAASRTPLALPLAFDQPGQVASRIVTHGI 361
 307 NVRVRFIPHEHALLPHVDAMVTNGYGGVNTALAFGVPLV-VARTEKEHYAARVSWSGA 365
 362 GKR-ASRFTTSHALARQIRSLTNTDYDQ 389
 366 GVHLKKRELSERDIRRAVAVL---DEPR 391

Search completed: February 29, 2004, 14:43:57
 Job time : 57.0584 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 24.9147 Seconds
(without alignments)
3937.172 Million cell updates/sec

title: US-09-941-947A-26
infect score: 1526
sequence: 1 LTVCAKHVHLTGISAEQLL.....HSTVQLFIQAWFKGLAAVS 303

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

minimum DB seq length: 0
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

database: SPTREMBL_25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1518	99.5	303	2 Q8GCS4	Q8GCS4 pantoea ste
2	1295.5	84.9	302	2 Q47842	Q47842 pantoea agg
3	1182.5	77.5	303	2 Q8VUJ8	Q8VUJ8 pantoea agg
4	490	32.1	359	2 Q93C19	Q93C19 xanthobacte
5	425	27.9	295	2 P94788	P94788 flavobacter
6	412.5	27.0	294	16 Q9K969	Q9K969 bacillus ha
7	399	26.1	316	2 Q9KIX4	Q9KIX4 bradyrhizob
8	391	25.6	299	16 Q8XX94	Q8XX94 ralstonia s
9	385.5	25.3	367	10 Q9SX26	Q9SX26 daucus caro
10	377.5	24.7	347	10 Q9ZU77	Q9ZU77 arabidopsis
11	376.5	24.7	293	16 Q8NWD6	Q8NWD6 staphylococ
12	376.5	24.7	357	10 Q9LIX0	Q9LIX0 arabidopsis
13	375.5	24.6	293	16 Q9RTX2	Q9RTX2 staphylococ
14	369.5	24.2	369	10 Q8LSC4	Q8LSC4 cistus inca
15	368	24.1	360	10 Q9LHR4	Q9LHR4 arabidopsis
16	365.5	24.0	303	16 Q7W7Q1	Q7W7Q1 bordetella

17	364	23.9	376	10	Q04046	O04046 arabidopsis
18	363.5	23.8	303	16	Q7WL38	Q7WL38 bordetella
19	363.5	23.8	303	16	Q7VW86	Q7VW86 bordetella
20	362.5	23.8	383	10	Q8WIR9	Q8WIR9 abiesgrand
21	361	23.7	372	10	Q9SLG2	Q9SLG2 arabidopsis
22	360.5	23.6	294	16	Q9AGM4	Q9AGM4 caulobacter
23	360	23.6	291	2	Q9RLH2	Q9RLH2 paracoccus
24	360	23.6	370	10	Q94ID7	Q94ID7 hevea bras
25	359.5	23.6	343	10	Q22043	Q22043 arabidopsis
26	359.5	23.6	356	10	Q81099	Q81099 helianthus
27	359.5	23.6	360	10	Q9LUB1	Q9LUB1 arabidopsis
28	359	23.5	305	16	Q98FG7	Q98FG7 rhizobium 1
29	359	23.5	371	10	Q8LAW5	Q8LAW5 arabidopsis
30	358.5	23.5	335	16	Q8UBX7	Q8UBX7 agrobacteri
31	358.5	23.5	336	10	Q9XIC0	Q9XIC0 arabidopsis
32	358.5	23.5	366	10	Q7XI92	Q7XI92 oryza sativ
33	357.5	23.4	304	16	Q8FYT3	Q8FYT3 brucella su
34	357.5	23.4	360	10	Q9LUD9	Q9LUD9 arabidopsis
35	357.5	23.4	393	10	Q9ZPM3	Q9ZPM3 taxus canad
36	356.5	23.4	304	16	Q8YJ16	Q8YJ16 bruceella me
37	355.5	23.3	360	10	Q8LBN9	Q8LBN9 arabidopsis
38	355	23.3	378	10	Q38917	Q38917 arabidopsis
39	354	23.2	302	16	P72683	P72683 synechocyst
40	353.5	23.2	360	10	Q8GY09	Q8GY09 arabidopsis
41	352.5	23.1	368	10	Q9SSU0	Q9SSU0 croton subl
42	352	23.1	309	16	Q8Z0B5	Q8Z0B5 anabaena sp
43	351.5	23.0	377	10	Q9SER3	Q9SER3 mentha pipe
44	349	22.9	309	16	Q7U881	Q7U881 synechococc
45	347	22.7	293	16	Q8CSG3	Q8CSG3 staphylococ

ALIGNMENTS

RESULT 1

Q8GCS4	PRELIMINARY;	PRT;	303 AA.
AC	Q8GCS4		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Geranylgeranyl pyrophosphate synthase.		
GN	CETE.		
OS	Pantoea stewartii.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Pantoea.		
OX	NCBI_TaxID=66269;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 8300;		
RA	deSouza M.L., Kollmann S.R., Schroeder W.A.;		
RT	"Carotenoid Biosynthesis (MO D2/079395 A2).";		
RL	Submitted (OCT-2002) to the EMBL/GenBank/DDBB databases.		
DR	EMBL; AY166713; AAN85596.1;		
DR	GO; GO:0008299; P:isoprenoid biosynthesis; IEA.		
DR	InterPro; IPR000092; Polyprenyl synth.		
DR	InterPro; IPR008949; Terpenoid synth.		
DR	pfam; PF00348; polyprenyl synth_1		
DR	PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.		
DR	PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.		
SQ	SEQUENCE 303 AA; 32430 MW; 5CBEF868FD7B432B CRC64;		

Query Match 99.5%; Score 1518; DB 2; Length 303;
Best Local Similarity 99.3%; Pred. No. 1.4e-114;
Matches 301; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	LTVCAKHVHLTGISAEQLLIDSRLLDQLLPVQGERDCVGAAMEGTLAPKRIPLML 60
Db	1	QYCAKHVHLTGISAEQLLIDSRLLDQLLPVQGERDCVGAAMEGTLAPKRIPLML 60
QY	61	LITARDLGCAISGGGLLDLACAVEMVHAASLILDDMPQMDAQMERGRPTIHTOYGEHVA 120
Db	61	LITARDLGCAISGGGLLDLACAVEMVHAASLILDDMPQMDAQMERGRPTIHTOYGEHVA 120

121	ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQVQGFKDLSEGDKPR	180
121		
121	ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQVQGFKDLSEGDKPR	180
121		
181	AILLTNQFKSTLPCASTQWASIAANASCEARENLRHSILDGOAFQLDDI-TDGM	240
181		
181	AILLTNQFKSTLPCASTQWASIAANASCEARENLRHSILDGOAFQLDDI-TDGM	240
181		
241	KDINQDAGKSTLVNLLGSGAVEERLRQHLASEHLSAACQNGHSTTQLFIQAFDK	300
241		
241	KDINQDAGKSTLVNLLGSGAVEERLRQHLASEHLSAACQNGHSTTQLFIQAFDK	300
241		
301	AVS 303	
301		
301	AVS 303	
301		
RESULT 2		
147842	PRELIMINARY; PRT; 302 AA.	
D	Q47842;	
D	Q47842;	
11	01-NOV-1996 (TrEMBLrel. 01, Created)	
11	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)	
11	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
11	CITE.	
11	CITE.	
11	Pantoea agglomerans.	
11	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
11	Enterobacteriaceae; Pantoea.	
11	NCBI_TaxID=549;	
11	[1]_	
11	PARTIAL SEQUENCE FROM N.A.	
11	MEDLINE=93371414; PubMed=8395826;	
11	Liu S.T.;	
11	"Carotenoid-biosynthesis genes as a genetic marker for the purpose of	
11	gene cloning.";	
11	Biochem. Biophys. Res. Commun. 195:259-263(1993).	
11	[2]	
11	SEQUENCE FROM N.A.	
11	MEDLINE=94236237; PubMed=8180698;	
11	To K.Y., Lai E.M., Lee D.Y., Lin T.P., Hung C.H., Chen C.L.,	
11	Chang Y.S., Liu S.T.;	
11	"Analysis of the gene cluster encoding carotenoid biosynthesis in	
11	Erwinia herbicola Ehcol3.";	
11	Microbiology 140:331-339(1994).	
11	ERMB; #90638; A8321260.1; --	
11	FIR; S52584; S52584.	
11	GO; GO:0008299; P:isoprenoid biosynthesis; IEA.	
11	InterPro; IPR000092; Polyphenyl_synth.	
11	InterPro; IPR008949; Terpenoid_synth.	
11	Pfam; PF00348; polyphenyl_synth; 1.	
11	PROSITE; PS00444; POLYPHENYL SYNTHET 2; 1.	
11	SEQUENCE 302 AA; 32626 MW; 9F443DAC894DFAFE CRC64;	
Query Match	84.9%; Score 1295.5; DB 2; Length 302;	
Best Local Similarity	86.1%; Pred. No. 1.4e-96;	
Matches 261; Conservative	14; Mismatches 27; Indels 1; Gaps 1	
QY	1 LTVCAKHVHLTGISAQELLADIISRLDGLLPVQGERDVCGAAMEGTLAPGKEIRPMLL	60
QY		
QY	1 MTVCAXKHVHTPSRAEQLLADIIRLQGLLPVEGERDFVGAAMEGALAPGKEIRPMLL	60
QY		
61	LITARDLGAISHGGLLDLACAVEMVHAASLILDDPCHDDAQMRRGSPTHITQYGHVA	120
61		
61	LITARDLGAISHGGLLDLACAVEMVHAASLILDDPCHDDAQMRRGSPTHITQYGHVA	120
61		
121	ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQVQGFKDLSEGDKPR	180
121		
121	ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQVQGFKDLSEGDKPR	180
121		
181	AILLTNQFKSTLPCASTQWASIAANASCEARENLRHSILDGOAFQLDDI-TDGM	240
181		

Db	181	AILMTNHFKTSTLFCASMQWASIVANASSERDYLHFRFSLDGLQAFQLLDDLTGCMADTG	240
Qy	241	KDINQDAGKSTLVNLGSGAVRERLQHLRLASEHLSAACNGHSTTQLFIOAWFDKKLA	300
Db	241	KDINQDEGKSTLVNLGSGRAVERLHSLHASEHLSAACNGHSTQQ-FIOAWFDKKLA	299
Qy	301	AVS 303	
Db	300	AVS 302	
RESULT 3			
Q8VUJ8	Q8VUJ8	PRELIMINARY;	PRT; 303 AA.
AC	Q8VUJ8;		
DT	01-MAR-2002 (TRENBUrel. 20, Created)		
DT	01-MAR-2002 (TRENBUrel. 20, Last sequence update)		
DT	01-OCT-2003 (TRENBUrel. 25, Last annotation update)		
DE	CrE protein.		
GN	CRTE.		
OS	Pantoea agglomerans pv. milletiae.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Pantoea.		
OX	NCBI_TaxID=18245;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Kamiunten H., Hirata R.;		
RT	"Isolation and characterization of carotenoid biosynthesis genes from		
RL	Pantoea agglomerans pv. milletiae wist 801."		
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB076662; BAB79500.1;		
DR	GO; GO:0008299; P:isoprenoid biosynthesis; IEA.		
DR	InterPro; IPR000092; Polyprenyl_synth.		
DR	InterPro; IPR008949; Terpenoid_synth.		
DR	Pfam; PF003348; polyprenyl_synth; 1.		
DR	PROSITE; PS00723; POLYPRENYL_SYNTHET 1; 1.		
DR	PROSITE; PS00444; POLYPRENYL_SYNTHET 2; 1.		
SQ	SEQUENCE 303 AA; 32504 MW; 6051833FDC0A5816D CRC64;		
Query Match 77.5%; Score 1182.5; DB 2; Length 303;			
Best Local Similarity 78.2%; Pred. No. 1.9e-87;			
Matches 237; Conservative 24; Mismatches 41; Indels 1; Gaps 1			
Qy	1	LTWCAKKEVHLTGISAECOLLADISRLDQLLPVQGRDCCVGAAMREGTILAPGKIRPMLL	60
Db	2	MTVCAEQHVFTHSDASLMDIEQRDLQLPVESERDLVGAMRUGALAPGKIRPLLL	61
Qy	61	LILTARDLGCATISHGGLDLACAVEMVHAASLIILDDWPCMDDAQMRGRPTTITQYGEHVA	120
Db	62	LLAARDLGCNATPAGLLDLACAVEMVHAASLIILDDWPCMDDAQLRGRPTTHCOYGEHVA	121
Qy	121	ILAAVALLSKAPGVTAEBGLTPIAKTVAVSELSTAIGMGLVQGFQDLSEGDKPRSAD	180
Db	122	ILAAVALLSKAPGVVAAAEGLTATARADAVELSHVAGMGLVQGFQDLSEGDKPRSAD	181
Qy	181	AILLTNOPKSTLFCASQTMASIAANASCSARENLRHFRSLDGLQAFQLLDDLTGCMADTG	240
Db	182	AILMTNHYKTSTLFCASQWASIVAAESGEAREQLHFRFSLNGLQAFQLLDDLTGCMADTG	241
Qy	241	KDINQDAGKSTLVNLGSGAVRERLQHLRLASEHLSAACNGHSTTQLFIOAWFDKKLA	300
Db	242	KDAAHQDGGKSTLVNLGSPQAVETRLDRHLRCASHLSACODGTATHH-FVQAWFEKLA	300
Qy	301	AVS 303	
Db	301	AVS 303	
RESULT 4			
ID	Q93CI9	PRELIMINARY;	PRT; 359 AA.
AC	Q93CI9;		
DT	01-DEC-2001 (T-EMBLrel. 19, Created)		

01-DEC-2001 {T-EMBLrel. 19, Last sequence update}
01-OCT-2003 {T-EMBLrel. 25, Last annotation update}
Geranylgeranyl diphosphate synthase.
CRTE.
Xanthobacter sp. (strain Py2).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Hyphomicrobiaceae; Xanthobacter.
NCBI_TaxID=78245;
[1]
SEQUENCE FROM N.A.
STRAIN=Py2;
Larsen R.A., Metcalf W.W.;
Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF408847; AL01998.1;
GO: 0008299; Pisprenoid biosynthesis; IEA.
InterPro; IPR000092; Polyprenyl synt.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synt; 1.
POSITIVE; PS00444; POLYPRENYL SYNTHET 2; 1.
SEQUENCE 359 AA; 37348 MW; AE33F9D22D4B727D CRC64;
Query Match 32.1%; Score 490; DB 2; Length 359;
Best Local Similarity 44.9%; Pred. No. 2.2e-31;
Matches 120; Conservative 41; Mismatches 88; Indels 18; Gaps 7
/ 19 LLADIISRLDQLPVGQERDCV-GAAMRETIAPGKRIRPMLLLLTARDLGCAGSHGGLL 77
71 LRAVIDRLGLVPPRAASHPAVLHAAMRHILLSPGKRRLPLLTLLAAAIQINA--SEHAVL 128
78 DLACAVEMVHAASLIILDDPCMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFGVIAE 137
129 DFGCALEMIHASSLIIVDDLPCMDDAAMRBAQPTTHVQYGEDVAVLASIGLLSRAFGVAAA 188
138 AEGLTPIAKTRAVSELSAIGMGLVQSGFKDI--SEGKPKPSADAILLTNQFKTSTLFC 195
189 APGVSEFARLEAVAILSMAGVSLGCGGYDDLRPSG---RSLNATEDVNRKTKGVLPFS 245
196 ASTQMASIAANASCEARENLRHRSFLDLGQAFQLDDLTDGMDT---GRDINQDACKSTL 252
246 AAVEIAGHVAADSTQGHLLKALAGHVGRAYQIILDDILDASASSASLGKDVGKDAHKATV 305
253 VNLGSGAVTEERLRQHURLASEHLSAA 279
306 IASLGA---PRAR--KLLSEHLAGA 325
RESULT 5
94788 PRELIMINARY; PRT; 295 AA.
C P94788
D P94788; 1997 {T-EMBLrel. 03, Created}
I 01-MAY-1997 {T-EMBLrel. 03, Last sequence update}
T 01-MAY-1997 {T-EMBLrel. 25, Last annotation update}
E Geranylgeranyl synthase.
N CRTE.
S Flavobacterium sp. ATCC 21588.
Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;
C Flavobacteriaceae; Flavobacterium.
X NCBI_TaxID=50286;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=R1534;
X MEDLINE=97186694; PubMed=9034310;
A Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,
van Loon A.P.;
T "Isolation and characterization of the carotenoid biosynthesis genes
of Flavobacterium sp. strain R1534.";
L Gene 185:35-41(1997).
R EMBL; U62808; AAC4848.1; --
R GO: 0008299; Pisprenoid biosynthesis; IEA.
R InterPro; IPR000092; Polyprenyl synt.
R InterPro; IPR008949; Terpenoid synth.
R Pfam; PF00348; polyprenyl synt; 1.
R POSITIVE; PS00444; POLYPRENYL SYNTHET 2; 1.

DR	PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
SQ	SEQUENCE 295 AA; 31332 MW; 849AC89EECEABC01 CRC64;
	Query Match 27.9%; Score 425; DB 2; Length 295;
	Best Local Similarity 39.6%; Pred. No. 3e-26;
	Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
QY	23 IDSRLLDQLLPVQGERDCV---CAAREGTLGPKRIKRLPMLLLLTARDLGCAISHGC--- 75
DB	13 VEIRLQAQ---ISQGFVSWAPLAGAASDAALSPKPRFRAVLMLVAE-----SSGGVCD 63
QY	76 -LLDLACAVEMVHAASLIILDDPCMDDAQMRRGRPTIHTOYGEHVAILAAVALLSKAFGV 134
DB	64 AMVDAAICAVEMVHAASLIIPDMFCMDARTRRGOPATHVAEGEGRAVLGIALITEAMRI 123
QY	135 IARAELGLTPAKTRAIVSELSATGMQGLVGQPFKLSEGDKPSADAILLTNQFKTSTLF 194
DB	124 LGARGATPPQARLVASMSRANGPVLCAGQDLNL--HAPDKAAGIEREQDLKTGVLF 180
QY	195 CASTOMASIANAASCARENLNHFSLDLGAQFOLLDDLTDGMTD---TGKDINQD-AGKS 250
DB	181 VAGLEMLSIIIKGLDKAEATEQLMAFGRLGRVFOSYDDLLDVIGDKASTGKTARDTAAPG 240
QY	251 TLVNLLGSQGAVERLRQHULASEHLSAACQNGHSTTOLF 290
DB	241 PKGCLMAVGQMGD-VAQHYRASQAQLDELME-----TRLF 274
RESULT 6	
ID	Q9K969 PRELIMINARY; PRT; 294 AA.
AC	Q9K969
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DI	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Geranyltranstransferase.
GN	BHZ781.
OS	Bacillus halodurans.
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX	NCBI_TaxID=86665;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C-125 / JCM 9153;
RX	MEDLINE=20512582; Pubmed=11058132;
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA	Horikoshi K.;
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus
RT	halodurans and genomic sequence comparison with Bacillus subtilis.";
RL	Nucleic Acids Res. 28:4317-4331(2000).
DR	EMBL; AP001516; BAB06500.1; -.
DR	PIR; E83997; E83997.
DR	GO; GO:0016740; P:transferase activity; IEA.
DR	GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR	InterPro; IPRO00092; Polyprenyl_synth.
DR	InterPro; IPRO08949; Terpenoid_synth.
DR	Pfam; PF00348; polyprenyl_synth; 1.
DR	PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR	PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW	Transferase; Complete proteome.
SQ	SEQUENCE 294 AA; 32653 MW; 3B2942868A62AD9E CRC64;
	Query Match 27.0%; Score 412.5; DB 16; Length 294;
	Best Local Similarity 37.6%; Pred. No. 3.1e-25;
	Matches 103; Conservative 47; Mismatches 113; Indels 11; Gaps 4;
QY	18 QLLADIDSRLDQLLPVQGER----DCVGAARMRGTTLAPGKRIRPMLLLLTTARDLGCASH 73
DB	7 QFLDSIKDIIIEERMPAHIERLNSPDLKSNMYSLKAGGKIRPALLLATWKSPQDISQ 66
QY	74 GGLLDLACAVEMVHAASLIILDDPCMDDAQMRRGRPTIHTOYGEHVAILAAVALLSKAFG 133
DB	67 G--IDLCAEMTHITSLIHDDLPMSDDDDIRGKPNTNHKVFGGAHALAGDALTYTFE 124

```
Y 134 VIAAEGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDADALLNQPSTL 193
b 125 IVAKMGVDPAKTLLEELARAAGPEGVGQVADIEGNGKLTVEGLEVIHHEKTCAL 184
Y 194 FCASQTAWASTANASCEARENLRHPSLDLGGAFOLLDLTDGKTD---TGKDNODAG-- 248
b 185 LSAFVAGARLADASQDITENIRFERELGLFQIKDDILDEVGDAAGKPGVSGDDGQ 244
Y 249 KSTLVNLGSGAVEERLRQHLRLASHLASAACON 282
b 245 KSTYPSLLTLEGAKELHLHTLLAKEVLYSVQVN 278

RESULT 7
9KIX4
D Q9KIX4 PRELIMINARY; PRT; 316 AA.
C Q9KIX4
Y 01-OCT-2000 (TReMBLrel. 15, Created)
Y 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
Y 01-OCT-2000 (TReMBLrel. 25, Last annotation update)
Y Geranylgeranyl synthase.
N CRTE.
S Bradyrhizobium sp. ORS278.
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
C Bradyrhizobiaceae; Bradyrhizobium.
X NCBI_TaxID=114615;
N
N SEQUENCE FROM N.A.
C STRAIN=ORS278;
X MEDLINE=20309720; PubMed=10851005;
X Hannibal L., Lorquin J., Angles d'Ortoli N., Garcia N.,
X Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;
X "Isolation and characterization of the canthaxanthin biosynthesis
X genes from the photosynthetic bacterium Bradyrhizobium sp. strain
X ORS278.";
X J. Bacteriol. 182:3850-3853 (2000).
D EMBL; AF218415; AAF78199.1; -.
D GO: 0008299; P: isoprenoid biosynthesis; IEA.
D InterPro; IPR000092; Polyprenyl synth.
D InterPro; IPR008949; Terpenoid synth.
D Pfam; PF00348; polyprenyl synth. 1.
D PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
D SEQUENCE 316 AA; 33713 MW; 9921ABD798A3FF16 CRC64;

Query Match 26.1%; Score 399; DB 2; Length 316;
Best Local Similarity 37.2%; Pred. No. 4.2e-24;
Matches 109; Conservative 38; Mismatches 106; Indels 40; Gaps 7;

Y 7 XHVLTGISBQLLADISRLDQLLPVQGRDCVGAARREGT-----LAPG 52
b 3 KPVDLTDTAA-----FETQLDRWRGRIGE--AVAEAMAFGTTVPAPLQAGMSHAVLAGG 54

Y 53 KRIRPMLLLTARDLGCAISHGGLLDACAVEMVHAASLDDMPDMDDAQMRGRGPTIH 112
b 55 KRYRGMVLVAGSLG--VPEQLSSAVAETTHAASLVVDLPDMDDARRRSQPAIE 112

Y 113 TOYGEHVAILAAVALLSKAFVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDL-- 170
b 113 VAFGEATAILSSIALIARAMEWARQRLSPASRSSIVDTLSHAIGFQALCGCGYDLYP 172

Y 171 ----SEGKPRSDADAILLTNQFTKSTLFCASQTAWASTANASCEARENLRHPSLDLGGAF 226
b 173 PYATEQD-----LHRYQRKTSALFVAFRCPTALLAEVDPETLLRIARACQRLGVAF 225

Y 227 QLLDPLTDGKTD---GKDNQDAGKSTLVNLGSGAVEERLRQHLRLASEHL 276
b 226 QIFDDLLDLTGDAHAGKVDQDGHGTVTLATILGPAAAEAADELAAVQKEL 278

RESULT 8
98XX94
D Q8XX94 PRELIMINARY; PRT; 299 AA.
C Q8XX94
Y 01-OCT-2000 (TReMBLrel. 15, Created)
Y 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
Y 01-OCT-2000 (TReMBLrel. 25, Last annotation update)
Y Geranylgeranyl synthase.
N CRTE.
S Bradyrhizobium sp. ORS278.
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
C Bradyrhizobiaceae; Bradyrhizobium.
X NCBI_TaxID=114615;
N
N SEQUENCE FROM N.A.
C STRAIN=ORS278;
X MEDLINE=20309720; PubMed=10851005;
X Hannibal L., Lorquin J., Angles d'Ortoli N., Garcia N.,
X Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;
X "Isolation and characterization of the canthaxanthin biosynthesis
X genes from the photosynthetic bacterium Bradyrhizobium sp. strain
X ORS278.";
X J. Bacteriol. 182:3850-3853 (2000).
D EMBL; AF218415; AAF78199.1; -.
D GO: 0008299; P: isoprenoid biosynthesis; IEA.
D InterPro; IPR000092; Polyprenyl synth.
D InterPro; IPR008949; Terpenoid synth.
D Pfam; PF00348; polyprenyl synth. 1.
D PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
D SEQUENCE 316 AA; 33713 MW; 9921ABD798A3FF16 CRC64;

Query Match 26.1%; Score 399; DB 16; Length 299;
Best Local Similarity 38.4%; Pred. No. 1.7e-23;
Matches 108; Conservative 43; Mismatches 104; Indels 26; Gaps 8;

Y 19 LEADIDSLDQLP---VOGRDCVGAARREGTLAPKRIKRPMLLLTARDLGCAISHGG 75
b 11 VVARTESALERALPGESVAPQR--LHAATYATIGAKRKVRP--LLAAAGALGEA--SPEA 66

Y 76 LLDLACAVEMVHAASLDDMPDMDDAQMRGRGPTIHQYGEHVAILAAVALLSKAFVI 135
b 67 LDGSCAVEMVHAASLDDMPDMDDDLLRGRFTVHRAVDEATALLVGDALQTOAFIAL 126

Y 136 AEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDADALLNQPSTLPC 195
b 127 AELGAVSPATRAGLVGELARASGSLGMAAGCAIDLQSGVALSQDALETHRMKTKALLR 186

Y 196 ASTOMASITANASCEARENLRHPSLDLGGAFOLLDLTDGKTD---GKDNQDAG 248
b 197 ASLRMGALCAGVNAALAEQVDAYAGAVGLAFQVDDILQVDTATLTGKTAGKDEAND-- 244

Y 249 KSTLVNLGSGAVEERLRQHLRLASEHLASAACQNGHSTTQL 289
b 245 KPTVVSILG-----LERARALADALHAAA--GAVAQL 275

RESULT 9
99SXZ6
D Q9SXZ6 PRELIMINARY; PRT; 367 AA.
C Q9SXZ6
Y 01-MAY-2000 (TReMBLrel. 13, Created)
Y 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
Y 01-OCT-2000 (TReMBLrel. 25, Last annotation update)
Y GPPP synthase.
N Daucus carota (Carrot).
S Daucus carota.
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C campanulids; Apiales; Apiaceae; Scandiceae; Daucinae;
C Daucus.
```

```
AC DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 25, Last annotation update)
DE Probable geranyltransferase (Farnesyl-diphosphate synthase)
DE protein (EC 2.5.1.10).
GN ISPA OR RSC2222 OR RS01377.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandelier M., Choisme N., Claudel-Renard C., Cunha S., Denange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin M., Schiex F.,
RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Welzenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
DR EMBL; AL646068; CAD15929.1; -.
DR GO: 0004337; F: geranyltransferase activity; IEA.
DR GO: 0016740; P: transferase activity; IEA.
DR GO: 0008299; P: isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth. 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 299 AA; 30911 MW; 7BF97371A1701382 CRC64;

Query Match 25.6%; Score 391; DB 16; Length 299;
Best Local Similarity 38.4%; Pred. No. 1.7e-23;
Matches 108; Conservative 43; Mismatches 104; Indels 26; Gaps 8;

Y 19 LEADIDSLDQLP---VOGRDCVGAARREGTLAPKRIKRPMLLLTARDLGCAISHGG 75
b 11 VVARTESALERALPGESVAPQR--LHAATYATIGAKRKVRP--LLAAAGALGEA--SPEA 66

Y 76 LLDLACAVEMVHAASLDDMPDMDDAQMRGRGPTIHQYGEHVAILAAVALLSKAFVI 135
b 67 LDGSCAVEMVHAASLDDMPDMDDDLLRGRFTVHRAVDEATALLVGDALQTOAFIAL 126

Y 136 AEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDADALLNQPSTLPC 195
b 127 AELGAVSPATRAGLVGELARASGSLGMAAGCAIDLQSGVALSQDALETHRMKTKALLR 186

Y 196 ASTOMASITANASCEARENLRHPSLDLGGAFOLLDLTDGKTD---GKDNQDAG 248
b 197 ASLRMGALCAGVNAALAEQVDAYAGAVGLAFQVDDILQVDTATLTGKTAGKDEAND-- 244

Y 249 KSTLVNLGSGAVEERLRQHLRLASEHLASAACQNGHSTTQL 289
b 245 KPTVVSILG-----LERARALADALHAAA--GAVAQL 275

RESULT 9
99SXZ6
D Q9SXZ6 PRELIMINARY; PRT; 367 AA.
C Q9SXZ6
Y 01-MAY-2000 (TReMBLrel. 13, Created)
Y 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
Y 01-OCT-2000 (TReMBLrel. 25, Last annotation update)
Y GPPP synthase.
N Daucus carota (Carrot).
S Daucus carota.
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C campanulids; Apiales; Apiaceae; Scandiceae; Daucinae;
C Daucus.
```

```

1 NCBI_TaxID=4039;
2 SEQUENCE FROM N.A.
3 TISSUE=Root;
4 "Daucus carota GDDP synthase 2.";
5 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
6 EMBL; AB027706; BAA78047.1;
7 GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
8 InterPro; IPR00092; Polyprenyl_synth.
9 Pfam; PF00348; polyprenyl_synth; 1.
10 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
11 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
12 SEQUENCE 367 AA; 40045 MW; 44DB284991A05E3 CRC64;

Query Match 25.3%; Score 385.5; DB 10; Length 367;
Best Local Similarity 37.2%; Pred. No. 6.3e-23;
Matches 99; Conservative 49; Mismatches 23; Indels 19; Gaps 7;

23 IDSRDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLLLTARDLGCAGSHGGLD 78
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 INKALDEAVPQ-KPAITHEAMRVSLLAGKRVPLCIAS-----CELV-GLSWAMP 139
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 LACAVEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGVIAEA 138
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 MACALEMTHTMSLIHDDLPCMDNDLRRGKPTNHKVFGEVAVLAGDALISPAPEHLAT 199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 -EGLTPIAKTRAVSELSTAIGMGLVQGVQKDLSEGDKPSADAILTNQKSTLFCAS 197
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 TDMVTPGRVFAIELGSAVSGVLVAGQINDICNQGKVDVLDLEVIHVKTSKLEAS 259
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 TOMASIAANASCEAPENLHRSLOLQAFQLDLDTGMDTGTGKINDQADG-----KST 251
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 VVCGALGGGVNIEVRKTKARCGILGFQVDDILD-VTKSSAEFGKTAGKOLIVTDKTT 318
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 LVNLLGSAVZEELRQHLRLASEHLS 277
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 YPKMLGRKRPATFELVAQAEELS 344
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
10077 PRELIMINARY; PRT; 347 AA.

01-MAY-1999 (TRENBLrel. 10, Created)
01-MAY-1999 (TRENBLrel. 10, Last sequence update)
01-OCT-2003 (TRENBLrel. 25, Last annotation update)
Putative geranylgeranyl pyrophosphate synthase.
AT2G18620
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
STRAIN=sv. Columbia;
MEDLINE=20083487; PubMed=10611797;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fuji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Sonerville C.R.,
Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
Salzberg S.L., Fraser C.M., Venter J.C.;
"Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
Nature 402:761-768(1999).
[2]
SEQUENCE FROM N.A.
STRAIN=sv. Columbia;
Lin X.;

```

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006135; RAD12206.1;
DR FIR; B84566; B84566.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR00092; Polyprenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00348; polyprenyl_synth; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR SEQUENCE 347 AA; 37390 MW; CC68CD1D22F8840 CRC64;

Query Match 24.7%; Score 377.5; DB 10; Length 347;
Best Local Similarity 37.6%; Pred. No. 2.6e-22;
Matches 89; Conservative 48; Mismatches 89; Indels 11; Gaps 4;

23 IDSRDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLLLTARDLGCAGSHGGLDLACA 82
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 LQSAVSLREPDK-----IHEAIRYSLLAGKRVPLCIAACELVGGESVA--LPAACA 122
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83 VEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGVIAEASGLT 142
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 VEMIHTMSLIHDDLPCMDNDLRRGKPTNHKVFGEVAVLAGDALISPAPEHLATSTAVS 182
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 PIATKTRAVSELSTAIGMGLVQGVQKDLSEGDKPS-----ADAILLTNQKSTLFCAS 199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 PARVTRAIAGELAKAIGSKLVAGQVVDLTSGMDNDVGLVLEFIHVHKTAVLELATV 242
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 MASIAANASCEAPENLHRSLOLQAFQLDLDTGMDTGTGKINDQADGKSTLVNLL 256
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
243 LGAIVGGSGDEEVEKLRERFARCGILGFQVDDILD-VTKSSEELGKTAGKOLIADKL 298
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 11
11077 PRELIMINARY; PRT; 293 AA.

01-OCT-2002 (TRENBLrel. 22, Created)
01-OCT-2002 (TRENBLrel. 22, Last sequence update)
01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE IspA protein.
GN ISPA OR MW1474.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
[1]
SEQUENCE FROM N.A.
MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamanoto K., Hiramatsu K.;
"Genome and virulence determinants of high virulence community-
acquired MRSA.";
Lancet 359:1819-1827(2002).
DR EMBL; AF004827; BAB95339.1;
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR00092; Polyprenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00348; polyprenyl_synth; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 293 AA; 32778 MW; 4A1920D061BB2CA7B CRC64;

Query Match 24.7%; Score 376.5; DB 16; Length 293;
Best Local Similarity 35.5%; Pred. No. 2.5e-22;
Matches 98; Conservative 44; Mismatches 119; Indels 15; Gaps 5;

11 LTGISAQQLADIDSRLDQLLPVQGERDCVGAAMREGTL-----APGKIRPMLLLTARD 66
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MTNLFMKNLIDVNNELS-----VAINKSWMDTQLEESMLYSLNAGGKRIRPVLILLTDS 56
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 LGCAISHGGLDLACAVEMVHAASLIIDMPCMDDAQMRGRPTHTQYGEHVAILAAVA 126
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

57 LNTYEGL--WKSAIALEMIHTVSLIHDDIAPMNDNDYRGKUTNHHVGVETAILAGDA 111
127 LLSKAFGVIAEAGLTPIAKTRAVSELSTAIGCSQGLVQGFQDLSECDKPRSADAILTN 186
115 LLTRAFELISSDDRLLTDEVKIKVLQRLSLASGHVGVGGQWLMDQSEGQPIDLETLEMIH 174
187 QFKSTLPECASTQWASIAANASCAARENHRFSLLDLCQAFQQLDDLTGQWTD---TGKDI 243
175 KTKTGALLTFVMSAADIATVDDATKSHLESYSYHLGWFMFOIKDDLDCYGDGAUKGKKV 234
244 NQDA--GKSTLVNLLGSAVEERLRQHLRLASEHLS 277
235 GSDLENNKSTVSLGKGADKLTYYHDAVDEL 270
RESULT 12
9LIAO PRELIMINARY; PRT; 357 AA.
ID Q9LIAO Q9LIAO:
YC 01-OCT-2000 (TREMBLrel. 15, Created)
YT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
YT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Geranylgeranyl pyrophosphate synthase.
DE Arabidopsis thaliana (Mouse-ear cress).
DC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
DC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
DC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_Taxid=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RL Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety Pl.,
RT TAC and BAC Clones";
RT DNA Res. 7:217-221 (2000).
DR EMBL; AF001309; BAB02589.1; -.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synt.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synt; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
SQ SSQUENCE 357 AA; 38502 MW; 41A40566CC8962F3 CRC64;
Query Match 24.7%; Score 376.5; DB 10; Length 357;
Best Local Similarity 37.1%; Pred. No. 3.3e-22;
Matches 103; Conservative 50; Mismatches 84; Indels 41; Gaps 10;
QY 23 ISRLDQLLPVGERDCVGAARBGTLAPGKRIKRMILL----LTARDLGAISHGGLD 78
DB 74 VSAALNVSVELQ-EPLTIQEAIVSYLLAGGKRVRELLCAACELVGGDEATAMS----- 132
QY 79 LACAVRMVHAASLILDMPCMDADWRGRPTIHTQYGHVAILAAVALISKAF----- 132
DB 127 AACAVENIHTSLIHDDLFQMDADALRGKPTNHHKEFGEDNAVLAGDALLALAEHMTFV 186
QY 133 --GVIAEAGLTPIAKTRAVSELSTAIGNQGLVQGFQDL-SEGDKPRSA--DAILTNQ 187
DB 187 SNGLVA-----PERMIRAVMELAKAIGTKGVAVQVTDLCSSQGLNPDDVGLERLEFIHL 240
QY 188 FKTSTLPECASTQWASIAANASCAARENHRFSLLDLCQAFQQLDDLTGQWTDGKINDQDA 244
DB 241 HKTAALAEAAVVGALMGGTETEEIKLRYARCGLLPQWDDILD-VTETSKELGTA 299
QY 248 GKST-----LVNLLGSGAVEERLRQHLRLASEHL 276

```

300 GKDWMAGKLTYPRLGILERSVREAKLR---REAABQL 334

Db

RESULT 13
Q99TX2 PRELIMINARY; PRT; 293 AA.

ID Q99TX2;
AC AC Q99TX2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE IsPa protein (Geranyltransferase homolog).
GN ISPA OR SAVI521 OR SAJ352.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi H., Kawara S., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kubura S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Ohshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi K., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003362; BAB57683.1; -;
DR EMBL; AP003134; BAB42614.1; -;
DR PTR; A85932; A85932.
GO; GO:0016740; F:transferase activity; IEA.
GO; GO:0008299; E:isoprenoid biosynthesis; IEA.
DR InterPro; IPRO00092; Polyprenyl_synth.
DR InterPro; IPRO08949; Terpenoid_synth.
DR Pfam; PF00348; polyprenyl_synt; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 293 AA, 32746 MW, 57AFEBBBA868F7D8 CRC64;

Query Match 24.6%; Score 375.5; DB 16; Length 293;
Best Local Similarity 35.5%; Pred.No.3e-22;
Matches 98; Conservative 44; Mismatches 119; Indels 15; Gaps 5

QY 11 LTGISAEOLLADIDSRLDOLLVPQGERDCVGAAAREGTL----APGRIRPMLLLLTARD 66
DB 1 MTNLPWKLIDEVNNLS-----VAINKSVMDTGLESSMLYSLNAGKRIRPVLLLTLDLS 56
QY 67 LGCAISHGLDLIDLACAVMVAASLIIDDMPCDDAQMERGRPTHTIQYGEHVAILAAVA 126
DB 57 LMTEYELG--VKSAIALSMITHYSLIHDDLPMANDNDYRGKLNTNHKVYGEWTALAGDA 114
QY 127 LLISKAGVTIAEAGITPTAKTRAVSELSTAIAGVLVQGOFKDSLGDKPRSDAILTN 186
DB 115 LUTKAPELISSDDRLLTDREVKKLVQLRSLASGEGVMGVGGMLDMQSEGPIDLETLEMIH 174
QY 187 QPKFTSTLFCASTQMASIAANASCARENLRFSLDLIGQAFOLLDLTDGWD---TGKDI 243
DB 175 KTKTKGALTTFVMSAADIANVDATKSHLSYSYHLGMFPQIKDDLLDCYGDEAKLGKKV 234
QY 244 NQDA--GKSTLVNLISGGAVERLRQHLRLASHLS 277
DB 235 GSDLNNKSYTVSLIGKGDAEDKLTYYRDAADVDEL 270

RESULT 14
Q8LSC4 PRELIMINARY; PRT; 369 AA.
ID Q8LSC4
AC Q8LSC4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)

RESULT 14	
Q8LSC4	
ID	Q8LS
AC	Q8LS
DT	01-0

01-OCT-2002 (TEMBLrel. 22, Last sequence update)
01-OCT-2003 (TEMBLrel. 25, Last annotation update)
Geranylgeranyl pyrophosphate synthase.

GGP2S2.
Cistus incanus subsp. creticus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Malvales; Cistaceae; Cistus.
NCBI_TaxID=191224;
[1]

SEQUENCE FROM N.A.
Pateraki I., Kaneellis A.K.;

"Cloning of Cistus creticus subsp. creticus geranylgeranyl
pyrophosphate synthase cDNA."

Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

EMBL; AF492023; AAM21639.1;
GO; GO:0008299; P:isoprenoid biosynthesis; IEA.

InterPro; IPR000092; Polyprenyl_synth.

InterPro; IPR008949; Terpenoid synth.

Pfam; PF00348; polyprenyl synth; 1.

PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.

PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.

SEQUENCE 369 AA; 39735 MW; 3AB04635AE021A3 CRC64;

Query Match 24.2%; Score 369.5; DB 10; Length 369;

Best Local Similarity 35.5%; Pred. No. 1.3e-21;

Matches 103; Conservative 49; Mismatches 111; Indels 27; Gaps 9;

17 EQLLADIDSRDLQPLVQGERDCVCAAREGTAPGKIRPMLLLTARDIGCAISHGGL 76

83 DQMAVSNOALESVSLR-EPKIHENRYSLLAGKKRFRPLLCIAACELVGGDVSA-- 139

77 LELACAVEMVHAASLIIDMPQMDQARRGRPTHTQYGEHVAIILAAVALLSKAFVIA 136

140 MPACAVEMVHTMSLIHDDLPCMNDDLRGKPTNHRKAFGEDIAVLACDALLSPAPEHVA 199

137 -EAGLTPIAKTRAVSELSTAIGMGIVQGFQDL-SEGDKPRSA--DAILLTNQKSTLTF 194

200 VSTVGASPDKIIVRAGVELAKAVGKGLVAGQVDTTSEGLNDVGLHLEYIHKHTAVLL 259

195 CASTOMASIAANASCEARENHRFSLDLGQAFQLDLDTD---GKTDGKIDNQD--AGK 249

260 EAAVGLGALGGGTDEEVERLEKFAICIGLLFQVVDLIDVTKSSVELGKTAGKDLVADK 319

250 STLNVLLG---SGAVEERLR-----QHILR-----LASEHLSAQQN 282

320 VTPEKLMGLEKSRFAEKLRDDAVQLRVFQVQKAAPLIALAHYIAYRQN 369

PRELIMINARY; PRT; 360 AA.

01-OCT-2000 (TEMBLrel. 15, Created)

01-OCT-2000 (TEMBLrel. 15, Last sequence update)

01-OCT-2003 (TEMBLrel. 25, Last annotation update)

Geranylgeranyl pyrophosphate synthase.

Arabidopsis thaliana (Mouse-ear cress).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

eurosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

[1]

SEQUENCE FROM N.A.

STRAIN=Columbia;

Nakamura Y.

"Structural Analysis of Arabidopsis thaliana Chromosome 3. III."

Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

EMBL; AP002033; BAB01936.1;

GO; GO:0008299; P:isoprenoid biosynthesis; IEA.

InterPro; IPR000092; Polyprenyl_synth.

InterPro; IPR008949; Terpenoid synth.

Pfam; PF00348; polyprenyl synth; 1.

DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
SQ SEQUENCE 360 AA; 38936 MW; 83A4F63A9E40CBC2 CRC64;

Query Match 24.1%; Score 368; DB 10; Length 360;
Best Local Similarity 37.6%; Pred. No. 1.6e-21;
Matches 97; Conservative 46; Mismatches 75; Indels 40; Gaps 9;

43 AMREGTLPAGKRIKIRPMLLL-----LTARDLGCAISHGGILLACAVEMVHAASLIIDDMPC 98

96 AVRYSLLAGKKRVRPLLCIAACELVGGDEATMS-----AACAVEMVHTSSLIHDDLPC 149

99 MDDAQMRRGRPTHTQYGEHVAIILAAVALLSKAF-----GVIAEAGLTPIAKTAV 150

150 MDDADLRGKPTNHRKVFGERMAVLADALLALAFEHMTVWSSGLVA-----PERMKSV 203

151 SELSTAIGMGIVQGFQDL-SEGDKPRSA--DAILLTNQKSTLTFCASTOMASIAANA 207

204 TELAKAIGTKGLVAGQVSDLSQGLNFDVGLERLEFIHLHKTAAALLEAAAVLGAIIGG 263

208 SCEARENHRFSLDLGQAFQLDLDTGMDTGTGKIDNDAKST-----LVNLLGS 258

264 TEEBIOKLRKYGRICIGLLFQVVDIID-VTSTBELGKTAGKDVMAKLTYPRLIGLRS 322

259 GAVEERLRQHLELASEHL 276

323 REVAEKLR---REAAEQ 337

Search completed: February 29, 2004, 14:50:49

Job time : 28.9147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:34:14 ; Search time 8.4386 Seconds
(without alignments)
3455.835 Million cell updates/sec

File: US-09-941-947A-26
Query score: 1526
Sequence: 1 LTVCAKXHVHVTGISAEOLL.....HSTTQLFIQAWFKKLAAYS 303

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 78:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1344.5	88.1	302	2	crTE protein - Erw
2	1295.5	84.9	302	2	crTE protein - Erw
3	752	49.3	308	2	phytoene synthase
4	412.5	27.0	294	2	geranyltransf
5	377.5	24.7	347	2	probable geranylge
6	375.5	24.6	293	2	hypothetical prote
7	372	24.4	369	2	farnesyltransf
8	369	24.2	371	2	geranylgeranyl pyr
9	364.5	23.9	366	2	farnesyltransf
10	364	23.9	376	2	geranylgeranyl pyr
11	361	23.7	372	2	probable geranylge
12	360.5	23.6	294	2	geranyltransf
13	360.5	23.5	297	2	geranyltransf
14	358.5	23.5	304	2	geranyltransf
15	358.5	23.5	335	2	geranyltransf
16	356.5	23.4	304	2	geranyltransf
17	355	23.3	378	2	geranylgeranyl pyr
18	354	23.2	302	2	geranylgeranyl pyr
19	352	23.1	309	2	geranylgeranyl dip
20	346.5	22.7	272	2	geranyltransf
21	345.5	22.6	316	2	farnesyltransf
22	345.5	22.6	357	2	farnesyltransf
23	339	22.2	294	2	geranyltransf
24	335.5	22.0	306	2	geranyltransf
25	332	21.8	299	2	geranyltransf
26	329	21.6	293	2	geranyltransf
27	328.5	21.5	298	2	geranyltransf
28	328	21.5	299	2	geranyltransf
29	328	21.5	299	2	geranyltransf

30 322 21.1 299 2 AG0554 geranyltransf
31 320 21.0 332 2 T10875 y4KU protein - Rhi
32 316 20.7 332 2 T40213 probable dimethyla
33 314 20.6 293 2 AC1245 geranyltransf
34 313.5 20.5 300 2 T06969 farnesyltransf
35 313 20.5 291 2 D82778 geranyltransf
36 313 20.5 326 2 T05674 farnesyltransf
37 309.5 20.3 295 2 C64123 geranyltransf
38 308 20.2 288 2 S49625 geranylgeranyl pyr
39 307 20.1 259 2 F81217 geranyltransf
40 307 20.1 288 2 T50750 farnesyltransf
41 305.5 20.0 291 2 F98007 geranyltransf
42 303.5 19.9 291 2 G95139 geranyltransf
43 301.5 19.8 295 2 F81139 geranyltransf
44 297.5 19.5 289 2 S04407 phytoene synthase
45 296 19.4 281 2 B81261 geranyltransf

ALIGNMENTS

RESULT 1

A37802
crTE protein - Erwinia uredovora
C:Species: Erwinia uredovora
C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
C:Accession: A37802
R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashir
J. Bacteriol. 172, 6704-6712, 1990
A:Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by funcitc
A:Reference number: A37802; MOID:9107214; PMID:2254247
A:Accession: A37802
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <MIS>
A:Cross-references: GB:D90087; NID:G216681; PIDN:BAAL4124.1; PID:G216682
C:Superfamily: geranyltransferase

Query Match 88.1%; Score 1344.5; DB 2; Length 302;
Best Local Similarity 88.4%; Pred. No. 1.7e-104;
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

Qy	1	LTVCAKXHVHVTGISAEOLLADIDSRDLQLLPVQGERDCVGAAMREGTTLAPGKRIRPMLL	60
Db	1	MTVCAXKHVHVTDAEQLLADIDRLDQLLPVEGERDVVGAAMREGALAPGKRIRPMLL	60
Qy	61	LLTARDLGCAISHGGLLDLACAVEMVHAAASLLDDMPQDDAQMRGRPTIHTQGEHVA	120
Db	61	LLTARDLGCAVSHDGLLDLACAVEMVHAAASLLDDMPQDDAKLRGRPTIHSYGEHVA	120
Qy	121	ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSAIGMGLVQGFQKLSGDKPRSD	180
Db	121	ILAAVALLSKAFGVIAADAGLTPLAKNRVSELSNAIGMGLVQGFQKLSGDKPRSAE	180
Qy	181	AILLTNQFKTSTLFCASQWASIAANASCEARENLFRLDGLQAFQQLDLDLTDGMDTGTG	240
Db	181	AILLTNHFKTSTLFCASQWASIVANASSEARDLHFRSLDGLQAFQQLDLDLTDGMDTGTG	240
Qy	241	KDINQDAGKSTLVNLLGSGAVEERLRQHLRLASEHLASAACQGHSTTFLQAWFKKLA	300
Db	241	KDSNQDAGKSTLVNLLGPRAVEERLRQHLRLASEHLASAACQGHGA-TQHFQAWFKKLA	299
Qy	301	AVS 303	
Db	300	AVS 302	

RESULT 2

S52584
crTE protein - Erwinia herbicola
C:Species: Erwinia herbicola
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 24-Nov-1999
C:Accession: S52584

Riim, Y.P.; Lai, B.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.
 Mol. Gen. Genet. 245, 417-423, 1994
 A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.
 A:Reference number: S52583; MUID:95107237; PMID:7808390
 A:Accession: S52584
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-302 <LIN>
 A:Cross-references: EXBL:M90698; NID:g148393; PIDN:AAA21260.1; PID:g148394
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
 C:Superfamily: geranyltransferase

Query Match 84.9%; Score 1295.5; DB 2; Length 302;
 Best Local Similarity 86.1%; Pred. No. 2e-100;
 Matches 261; Conservative 14; Mismatches 27; Indels 1; Gaps 1;

QY 1 LTVCAKXVHLTGSAEQLADISRLDQLPVQGERDCVGAAREGTLAPGKRIKRPML 60
 DB 1 MIVCAKXVPTPSAAQLLADIRLEQLPVQGERDPVGAAREGTLAPGKRIKRPML 60
 QY 61 LITARDLGCAISHGGLDLACAVEMVHAASLILDDMPGQKLSAECQHGSHSTTQFIQAFDCKLA 120
 DB 61 LITARDLGCAVSHGGLDLACAVEMVHAASLILDDMPGQKLSAECQHGSHSTTQFIQAFDCKLA 120
 QY 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFYDLSEGDKPRSA 180
 DB 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFYDLSEGDKPRSA 180
 QY 181 AILLTNOFTSTLFCASOMASIVANASERRDYLHRSFSLDGLQAFQLDLDTGMDTG 240
 DB 181 AILLTNEFTSTLFCASOMASIVANASERRDYLHRSFSLDGLQAFQLDLDTGMDTG 240
 QY 241 KQINDQKSTLVNLLGSGAVERLRQHLRLASEHLSAACQHGSHSTTQFIQAFDCKLA 300
 DB 241 KQINDQKSTLVNLLGSGAVERLRQHLRLASEHLSAACQHGSHSTTQFIQAFDCKLA 300
 QY 301 AVS 303
 DB 300 AVS 302

RESULT 3
 C39273
 phytoene synthase - Erwinia herbicola
 C:Species: Erwinia herbicola
 C:Date: 18-Oct-1991 #sequence_revision 03-Feb-1994 #text_change 15-Oct-1999
 A:Accession: C39273; C33120
 R:Armstrong, G.A.; Alberti, M.; Hearst, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990
 A:Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in non
 A:Reference number: A39273; MUID:91088634; PMID:2263648
 A:Accession: C39273
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-308 <ARM>
 A:Cross-references: GB:M38424
 A>Note: the authors translated the codon CTC for residue 187 as Ala, ATT for residue 191
 e 211 as Ala
 A>Note: the authors inserted as Arg after residue 261 in figure 3
 R:Armstrong, G.A.; Alberti, M.; Hearst, J.E.
 submitted to the Protein Sequence Database, September 1990
 A:Reference number: A33120
 A:Accession: C33120
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-261,263-308 <AR2>
 A:Experimental source: strain Eho10, ATCC 39368
 C:Genetics:
 A:Gene: crtE
 C:Superfamily: geranyltransferase
 C:Keywords: carotenoid biosynthesis

Query Match 49.3%; Score 752; DB 2; Length 308;

Best Local Similarity 55.8%; Pred. No. 4e-55;
 Matches 159; Conservative 40; Mismatches 84; Indels 2; Gaps 2;

QY 17 EOLLADIISRLDQLPVQGERDCVGAAREGTLAPGKRIKRPMLLTARDLGCAISHGGL 76
 DB 16 EVMROSIDHLAGLIPETDSQDIVSLAVEGVMAPGKRIKRPMLLAARDLRQGSNPTL 75
 QY 77 DLACAVEMVHAASLILDDMPGQKLSAECQHGSHSTTQFIQAFDCKLA 136
 DB 76 DLACAVELTTPASIMLDDMPGQKLSAECQHGSHSTTQFIQAFDCKLA 135
 QY 137 EAGELTPIAKTRAVSELSTAIGMGLVQGFYDLSEGDKPRSAAILTNOFTSTLFC 196
 DB 136 ATGDLPGERRACAVNELSTAVGQGLVLGQERDLNDAALDRTPDAILSTNHLKTLFSA 195
 QY 197 STOMASIAANASCEARENHRSFSLDGLQAFQLDLDTGMDTGKINDQAGKSTLVNLL 256
 DB 196 MEOIVAIASASSPSTRETTHAFALDFQAFQLDLDDHPETGKDRNKDAGKSTLVNRL 255
 QY 257 GSGAV-EERLRQHLRLASEHLSAACQHGSHSTTQFIQAFDCKLA 300
 DB 256 GADAAROKLREHIDSADKHLTFACPGQAIRO-FMELWFGHLLA 299

RESULT 4
 E83997
 geranyltransferase BH2781 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-May-2002
 C:Accession: E83997
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hir
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: AB3650; MUID:20512582; PMID:11058132
 A:Accession: E83997
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-294 <STO>
 A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06500.1; GSPDB:GNO
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2781
 C:Superfamily: geranyltransferase

Query Match 27.0%; Score 412.5; DB 2; Length 294;
 Best Local Similarity 37.6%; Pred. No. 7.5e-27;
 Matches 103; Conservative 47; Mismatches 113; Indels 11; Gaps 4;

QY 18 QLLADIDSRLDQLPVQGER----DCVGAAREGTLAPGKRIKRPMLLTARDLGCAISH 73
 DB 7 QFLDEIKDIIERRPAHIERLNSPDMKNSMLYSKAGGKRIRPALLATMKSFKQDISQ 66
 QY 74 GGLDLACAVEMVHAASLILDDMPGQKLSAECQHGSHSTTQFIQAFDCKLA 133
 DB 67 G-IDLACAEIEMHTYSLIHDDLPSMDDDDIRRGKPTNHKVFGEAHAILAGDALLTYSFE 124
 QY 134 VIAEAGLTPIAKTRAVSELSTAIGMGLVQGFYDLSEGDKPRSAAILTNOFTSTL 193
 DB 125 IVAKKMGVDPKATLCLEELARAAGPGVGGVADIEGENOKLTVEGLEIYTHHRTKAL 184
 QY 194 FCASOMASIAANASCEARENHRSFSLDGLQAFQLDLDTGMDT---TGKINDQAG-- 248
 DB 185 LSPFAVAGARLADASEQDIENIRFRSRELGLLFOIKDDILDDVEGQAAIGKPVGSDGNG 244
 QY 249 KSTLVNLLGSGAVERLRQHLRLASEHLSAACON 282
 DB 245 KSTYPSLLTLEGAKEKHLHTLLAKEVLYSVQMN 278

RESULT 5
 E84566
 probable geranylgeranyl pyrophosphate synthase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

```
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
;Accession: E84566
;Link, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
;Lin, X.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
;Koo, H.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
;ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Status: preliminary
;Molecule type: DNA
;Residues: 1-347 <STO>
;Cross-references: GB:AE002093; NID:94217998; PIDN:RAD12206.1; GSPDB:GN00139
;Genetics:
;Gene: At2g18620
;Map position: 2
;Superfamily: geranyltransferase

Query Match 24.7%; Score 377.5; DB 2; Length 347;
Best Local Similarity 37.6%; Pred. No. 7.6e-24;
Matches 89; Conservative 48; Mismatches 89; Indels 11; Gaps 4;

Y 23 IDSRLDQLLPVQGERDCVGAAMEEGTLAPGKRIRPMLLLLTARDLGCAISHGGLDLACA 82
b 70 LGSVSLRLEPIK-----IHEAIRYSELARGKVRPVLCTAACELVGGEESVA--LPAACA 122
Y 83 VEMVHAASLLIDMPCDDAQMGRGPTTIHQYGEHVALLAAVALLSKAFGVIAEAGLT 142
b 123 VEMHTMSLHDDLPCDDNDLARGPTNHKIVGEDVAVLAGSLAPAFEHVNSTAGV 182
Y 143 PIAKTRAVSELSTAIGMQLVQGVQFQKLS-EGDKPRSDAILLTNQFTSTLFCASTQ 199
b 183 PARVRAIGELAKIGSKGLVAGVVDVLSGGWDQNDVGLVLEFIHVHKTAVLEAATV 242
Y 200 MASTANASCEARENLRHFSLDLQGAQFOLLDDLTGMDTGKIDNODAGKSTLVNLL 256
b 243 LGAIVGGSDSEVEKLRERFARCIIGLLFQVDDILD-VTKSSEELGKTAGKDLVADKL 298

RESULT 6
Accession: A8932
Species: Staphylococcus aureus
;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
;Accession: A8932
;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguro,
;na, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
;C.; Shiba, T.; Rattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
;ancet 357, 1225-1240, 2001
;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
;Reference number: A89758; MUID:21311952; PMID:11418146
;Accession: A8932
;Status: preliminary
;Molecule type: DNA
;Residues: 1-293 <KUR>
;Cross-references: GB:BA000018; PID:gl3701320; PIDN:BA842614.1; GSPDB:GN00149
;Experimental source: strain N315
;Genetics:
;Gene: ispA

Query Match 24.6%; Score 375.5; DB 2; Length 293;
Best Local Similarity 35.5%; Pred. No. 9e-24;
Matches 98; Conservative 46; Mismatches 119; Indels 15; Gaps 5;

Y 11 LTGISABQLADISRLDQLLPVQGERDCVGAAMEEGTL----APGKRIRPMLLLLTARD 66
b 1 MTNLPMNKLIDEVNNEL-----VAINKSVMDTQLEESMLYSINAGGKRIRPVLMLTLDS 56
Y 67 LGCAISGGGLDLACAVEMVHAASLLIDMPCDDAQMGRGPTTIHQYGEHVALLAAVA 126
b 57 LNTYEYELG--VKSAIALEMHTHTYSLHDDLPEAMNDYRGRKLTNKHVYGETAILAGDA 114
Y 127 LLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGVQFQKLS-EGDKPRSDAILLTN 186
```

```
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
;Accession: E84566
;Link, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
;Lin, X.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
;Koo, H.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
;ature 402, 761-768, 1999
;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
;Reference number: A84420; MUID:20083487; PMID:10617197
;Status: preliminary
;Molecule type: DNA
;Residues: 1-347 <STO>
;Cross-references: GB:AE002093; NID:94217998; PIDN:RAD12206.1; GSPDB:GN00139
;Genetics:
;Gene: At2g18620
;Map position: 2
;Superfamily: geranyltransferase

Query Match 24.7%; Score 377.5; DB 2; Length 347;
Best Local Similarity 37.6%; Pred. No. 7.6e-24;
Matches 89; Conservative 48; Mismatches 89; Indels 11; Gaps 4;

Y 23 IDSRLDQLLPVQGERDCVGAAMEEGTLAPGKRIRPMLLLLTARDLGCAISHGGLDLACA 82
b 70 LGSVSLRLEPIK-----IHEAIRYSELARGKVRPVLCTAACELVGGEESVA--LPAACA 122
Y 83 VEMVHAASLLIDMPCDDAQMGRGPTTIHQYGEHVALLAAVALLSKAFGVIAEAGLT 142
b 123 VEMHTMSLHDDLPCDDNDLARGPTNHKIVGEDVAVLAGSLAPAFEHVNSTAGV 182
Y 143 PIAKTRAVSELSTAIGMQLVQGVQFQKLS-EGDKPRSDAILLTNQFTSTLFCASTQ 199
b 183 PARVRAIGELAKIGSKGLVAGVVDVLSGGWDQNDVGLVLEFIHVHKTAVLEAATV 242
Y 200 MASTANASCEARENLRHFSLDLQGAQFOLLDDLTGMDTGKIDNODAGKSTLVNLL 256
b 243 LGAIVGGSDSEVEKLRERFARCIIGLLFQVDDILD-VTKSSEELGKTAGKDLVADKL 298

RESULT 7
Accession: S53722
Species: Capsicum annuum (pepper)
;Date: 15-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
;Accession: S53722
;R;Badillo, A.; Steppuhn, J.; Derruere, J.; Camara, B.; Kuntz, M.
;Plant Mol. Biol. 27, 425-428, 1995
;A;Title: Structure of a functional geranylgeranyl pyrophosphate synthase gene from Caps1
;Reference number: S53722; MUID:95195169; PMID:7886631
;Accession: S53722
;A;Status: nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-369 <BAD>
;Cross-references: EMBL:X80267; NID:9643093; PIDN:CAA56554.1; PID:9643094
;A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
;C;Genetics:
;A;Gene: GGPPS
;C;Superfamily: geranyltransferase
;C;Keywords: carotenoid biosynthesis; chloroplast; chromoplast; transferase

Query Match 24.4%; Score 372; DB 2; Length 369;
Best Local Similarity 34.5%; Pred. No. 2.4e-23;
Matches 99; Conservative 52; Mismatches 108; Indels 28; Gaps 7;

Y 23 IDSRLDQLLPVQGERDCVGAAMEEGTLAPGKRIRPMLLLLTARDLGCAISHGGLDLACA 82
b 89 VNKALDEAIIIVK-EPHVIHEAMRYSLLAGGKRVPMCLAAACELVGG--NQENAMAACA 145
Y 83 VEMVHAASLLIDMPCDDAQMGRGPTTIHQYGEHVALLAAVALLSKAF-GVIAEAGL 141
b 146 VEMHTMSLHDDLPCDDNDLARGPTNHKIVGEDVAVLAGSLAPAFEHVNSTAGV 205
Y 142 TPIAKTRAVSELSTAIGMQLVQGVQFQKLS-EGDKPRSDAILLTNQFTSTLFCASTQ 200
b 206 TPSRIIVGVAELAKSIGTEGLVAGVADIKCTGNASVSLTLEFIHVHKTAALESSVWL 265
Y 201 ASIANNASCEARENLRHFSLDLQGAQFOLLDDLTGMDTGKIDNODAGKSTLVN----- 254
b 266 GAILGGGTNVEVEKLRERFARCIIGLLFQVDDILD-VTKSSEELGKTAGKDLVADKLTYPK 324
Y 255 LLGSGVAEERLRHRLHRLASEHLSAAACQNGHSTTQTFQIAWFDKXKLA 301
b 325 LLGLEKAEPAELNREAKQQLG-----FDSKAA 355

RESULT 8
Accession: F85434
Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 24-Aug-2001
;Accession: F85434
;R;Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
;Nature 402, 769-777, 1999
;A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
;Reference number: A85001; MUID:20083488; PMID:10617198
;Accession: F85434
;Status: preliminary
;Molecule type: DNA
```

```

;Residues: 1-371 <STO>
;Cross-references: GB:NC_001268; NID:g7270630; PIDN:CAB80347.1; GSPDB:GN00140
;Genetics:
;Gene: AY936810
;Map position: 4
;Superfamily: geranyltransferase

Query Match      24.2%; Score 369; DB 2; Length 371;
Best Local Similarity 38.8%; Pred. No. 4.2e-23;
Matches 93; Conservative 46; Mismatches 91; Indels 10; Gaps 6;

y 23 IDSRLDQLLPVQGERDCVGAAMREGTLAPGRIRPMLLLTARDLGCALSHGGLDLACA 82
b 87 VNKALDSAVPLR-EPLKHEAMRYSLAGGKVRPVLCAACELVGGEBSTA-MPAACA 143
y 83 VEMVHAASLILDDMPCMDAQMGRGRTHTQYGEHVAILAAVALLSKAFGVIAEA---E 139
b 144 VEMHTMSLIHDDLPCMDNDLRRKPTNEKVFGEVDVAVLAGDALLSFAPEHLASATSSD 203
y 140 GLTPPIAKTRAVSELSTAIGMGLVQGFQKDL-SGDKPRSA--DAILLTNQFTSTLPCA 196
b 204 VVSPVRVAVGELAKALGTGLVAGQVVDISSEGLDNDVGLHLEFTHLHKTAALEA 263
y 197 STQMASIAANASCRARENHLPFSLDGOAFOLLDDLTGDMTDTGKINODAGKSTLWLL 256
b 264 SAVLGAIVGGSDDEIERLRKFAKIGLGFQVVDIILD-VTKSKELGKTAGKDLADKL 322

RESULT 9
10452
;Arnyltransferase (EC 2.5.1.29) precursor, chloroplast - white mustard
;Alternate names: geranylgeranyl-diphosphate synthase; geranylgeranyl-pyrophosphate syn
;Species: Sinapis alba (white mustard)
;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
;Accession: T10452
;Author: Bonk, M.; Hoffmann, B.; von Lintig, J.; Schledz, M.; Al-Babili, S.; Hobeika, E.; Klein
;Title: Chloroplast export of four carotenoid biosynthetic enzymes in vitro reveals dif
;Reference number: Z17023; MUID:97433278; PMID:9288918
;Accession: T10452
;Status: preliminary; translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-366 <BON>
;Cross-references: EMBL:X98795; NID:gl419757; PIDN:CAA67330.1; PID:gl419758
;Genetics:
;Gene: GGPS
;Genome: nuclear
;Superfamily: geranyltransferase
;Keywords: carotenoid biosynthesis; chloroplast; transferase

Query Match      23.9%; Score 364.5; DB 2; Length 366;
Best Local Similarity 39.5%; Pred. No. 9.9e-23;
Matches 96; Conservative 42; Mismatches 88; Indels 17; Gaps 8;

y 23 IDSRLDQLLPVQGERDCVGAAMREGTLAPGRIRPMLLLTARDLGCALSHGGLDLACA 80
b 83 VNKALDSAVPLR-EPLKHEAMRYSLAGGKVRPVLCAACELV-GGEBSLAMP 135
y 81 --CAVEMVHAASLILDDMPCMDAQMGRGRTHTQYGEHVAILAAVALLSKAFGVIAEA 138
b 136 ARCAVEMHTMSLIHDDLPCMDNDLRRKPTNEKVFGEVDVAVLAGDALLSFAPEHLASA 195
y 139 EG--LTPPIAKTRAVSELSTAIGMGLVQGFQKDL-SGDKPRSA--DAILLTNQFTSTL 193
b 196 TSSEVSPARVAVGELAKALGTGLVAGQVVDISSEGLDNNVGLHLEFTHLHKTAAAL 255
y 194 FCASQMASIAANASCRARENHLPFSLDGOAFOLLDDLTGDMTDTGKINODAGKSTLV 253
b 256 LEASAVTGGIIGGSDDEIERLRKFAKIGLGFQVVDIILD-VTKSSQELGKTAGKDLIA 314
y 254 NLL 256
b 315 DKL 317

```

RESULT 10

```

T02429
;Arnyltransferase (EC 2.5.1.1) - Arabidopsis thaliana
;Alternate names: pregeranylgeranyl pyrophosphate synthase; protein F27L4.2; protein P;
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Feb-2001
;Accession: T02429; T02430; B84629
;Author: Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
;Title: The EMBL Data Library, May 1998
;Description: Arabidopsis thaliana chromosome II BAC F27L4 genomic sequence.
;Reference number: Z14658
;Accession: T02429
;Status: translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-376 <ROU>
;Cross-references: EMBL:AC004482; NID:g3152602; PIDN:AAC17083.1; PID:g3152604
;Experimental source: cultivar Columbia
;Accession: T02430
;Status: translated from GB/EMBL/DDBJ
;Molecule type: DNA
;Residues: 1-336 <ROW>
;Cross-references: EMBL:AC004482; NID:g3152602; PIDN:AAC17084.1; PID:g3152605
;Experimental source: cultivar Columbia
;Author: R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
;Title: Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
;Author: euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, A.;
;Title: Nature 402, 761-768, 1999
;Reference number: A84420; MUID:20083487; PMID:10617197
;Accession: B84629
;Status: preliminary
;Molecule type: DNA
;Residues: 1-376 <STO>
;Cross-references: GB:AE002093; NID:g3152604; PIDN:AAC17083.1; GSPDB:GN00139
;Genetics:
;Gene: At2g23800; F27L4.2; F27L4.3
;Map position: 2
;Introns: 184/3
;Superfamily: Geranyltransferase
;Keywords: transferase

Query Match      23.9%; Score 364; DB 2; Length 376;
Best Local Similarity 38.1%; Pred. No. 1.1e-22;
Matches 102; Conservative 45; Mismatches 97; Indels 24; Gaps 9;

y 23 IDSRLDQLLPVQGERDCVGAAMREGTLAPGRIRPMLLLTARDLGCALSHGGLDLACA 82
b 96 VNKALDSAVPLR-EPLKHEAMRYSLAGGKVRPVLCAACELVGG--QENAAAPACA 152
y 83 VEMVHAASLILDDMPCMDAQMGRGRTHTQYGEHVAILAAVALLSKAFGVIAEGLT 142
b 153 VEMHTMSLIKDDLPMDNDLRRKPTNEKVFGEVDVAVLAGDALLSFAPEHMTAAE--- 209
y 143 PIAKTR---AVSELSTAIGMGLVQGFQKDL-SGDKPRSA--DAILLTNQFTSTLPCA 196
b 210 -ISSERMVAVRELARSIGTRGVAGQAMDISEGLDNEVGLHLEFTHVHKTAVLET 268
y 197 STQMASIAANASCRARENHLPFSLDGOAFOLLDDLTGDMTDTGKINODAGKSTLV 248
b 269 AAVLGAIVGGSDDEIERLRKFAKIGLGFQVVDIILD-VTKSKELGKTAGKDLIA 325
y 249 KSTLVNLLGSAVPELRHLEFTHVHKTAVLET 276
b 326 KLTYPKLGLEKSKFVKELTKDAQHL 353

RESULT 11
G84566
;Probable geranylgeranyl pyrophosphate synthase [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002

```

Accession: G84566
Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Varakan, S.B.; Umayam, L.; Tallon, L.;
Bass, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, C.M.;
Genome Science Center, J. Craig Venter Institute, 1999
Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
Reference number: A84420; MUID:20083487; PMID:10617197
Accession: G84566
Status: preliminary
Molecule type: DNA
Residues: 1-372 <STO>
Cross-references: GB:AE002093; NID:g4185130; PIDN:AA08933.1; GSPDB:GN00139
Genetics:
Gene: At2g18640
Map position: 2
Superfamily: geranyltransferase

Query Match 23.7%; Score 361; DB 2; Length 372;
Best Local Similarity 37.4%; Pred. No. 2e-22;
Matches 99; Conservative 47; Mismatches 101; Indels 18; Gaps 8;

23 IDSRDLQLPVGQEDCVGAAMRGSTLAPGKIRPMLLLTARDLGAISHGG---LLDL 79
29 IDSRDLQLPVGQEDCVGAAMRGSTLAPGKIRPMLLLTARDLGAISHGG---LLDL 79
92 INRALDEAIPIT-BELNTHKAVYAILAGGKVRPIIQL-----ACELVGGGERIAQA 145
98 INRALDEAIPIT-BELNTHKAVYAILAGGKVRPIIQL-----ACELVGGGERIAQA 145
80 ACAYEMVHAASLIILDDMPCMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFVIAEAE 139
86 ACAYEMVHAASLIILDDMPCMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAFVIAEAE 139
146 ACAYEMVHTMSLIKDDLPMDNDLDRGKPTTHKVFGEVAILSGGALLALAEHLTEAD 205
152 ACAYEMVHTMSLIKDDLPMDNDLDRGKPTTHKVFGEVAILSGGALLALAEHLTEAD 205
140 GLTPIAKTRAVSELSTAIGMGLVQGGKDL-SGGDKPRSA--DAIILLTQFKTSTILFCA 196
146 GLTPIAKTRAVSELSTAIGMGLVQGGKDL-SGGDKPRSA--DAIILLTQFKTSTILFCA 196
206 -VSSKKQWRAVKELAKSITGRLVAGQAKDLSSEGLEQNDVGLLEDLEYIHHVTKTGSLLA 264
212 -VSSKKQWRAVKELAKSITGRLVAGQAKDLSSEGLEQNDVGLLEDLEYIHHVTKTGSLLA 264
197 STQMASIAANASCEARENLEHFSLDLGAQFOLLDDLT---GMTDTGCKINQD--AGKST 251
203 STQMASIAANASCEARENLEHFSLDLGAQFOLLDDLT---GMTDTGCKINQD--AGKST 251
265 SAVIGAVGGGTEKEIEKVRFPACIGLLFQVDDILDETKSSEELCKTAGKQKVGAKLT 324
271 SAVIGAVGGGTEKEIEKVRFPACIGLLFQVDDILDETKSSEELCKTAGKQKVGAKLT 324
325 YPKVIGVEKSEFEVKKLKDAREHL 349
331 YPKVIGVEKSEFEVKKLKDAREHL 349

522 LYNLLGSGAVERLRFQHLRLASEHL 276
528 LYNLLGSGAVERLRFQHLRLASEHL 276
325 YPKVIGVEKSEFEVKKLKDAREHL 349
331 YPKVIGVEKSEFEVKKLKDAREHL 349

RESULT 12
87505
Species: Caulobacter crescentus
Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
Accession: D87505
Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eissen, J.; Heidelberg, J.;
Kaul, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.;
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
Title: Complete Genome Sequence of *Caulobacter crescentus*.
Reference number: A87249; MUID:21173698; PMID:11259647
Accession: D87505
Status: preliminary
Molecule type: DNA
Residues: 1-294 <STO>
Cross-references: GB:AE005673; NID:gi3423548; PIDN:AAK24040.1; GSPDB:GN00148
Genetics:
Gene: CC2069

Query Match 23.6%; Score 360.5; DB 2; Length 294;
Best Local Similarity 36.5%; Pred. No. 1.6e-22;
Matches 105; Conservative 39; Mismatches 101; Indels 43; Gaps 11;

21 ADTDS-RLDQLAP-VQGERDCVGAAMRGSTLAPGKIRPMLLLTARDLGAISHGGLLD 78
27 ADTDS-RLDQLAP-VQGERDCVGAAMRGSTLAPGKIRPMLLLTARDLGAISHGGLLD 78
12 ADIVTALDELPRADGPESRLTEAMRYALGPKRLRPFPALET-----GKMF 61
18 ADIVTALDELPRADGPESRLTEAMRYALGPKRLRPFPALET-----GKMF 61
79 L-----ACAYEMVHAASLIILDDMPCMDDAQMRGRPTIHTQYGEHVAILAAVALLSK 130
85 L-----ACAYEMVHAASLIILDDMPCMDDAQMRGRPTIHTQYGEHVAILAAVALLSK 130
62 LPERPVLRAACALECTHAYSLVHEDLPAMDNDVVRGPTVHKQYDEATLADGALQTA 121
68 LPERPVLRAACALECTHAYSLVHEDLPAMDNDVVRGPTVHKQYDEATLADGALQTA 121

```

QY 131 AFGVIAEAGLTP-IAKTRAVSELSTAIGMOGLVQGGFKLSEGDKPSADAILTNQFK 189
DB   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 122 AFEIWAHPDTHDDGHVRSSELVRKLAIASGARGCGGQWIDLL-GVRD-DLGAVARMQRLK 179
DB   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 190 TSTLFCASTOMASIAANASCEARENLHRPSLDLGOAFOLLDDLT-D-----GMTDTGK 241
DB   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 180 TGLAIAPAFELPIIIPARAKESRSALMAFADLGLAYQIVDDILDAGDEETLGAAGGK 239
DB   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 242 DINODAGKSTLVNLGSGAVBERL-----RQHRL---ASEHLSAA 279
DB   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 240 DAAK--GRANVTVLLGLDAAXERVTLAEQTRSHLEIFGERAEHLRAS 285
DB   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 13
JX0257
geranyltranstransferase (BC 2.5.1.10) - Bacillus stearothermophilus
N:Alternate names: farnesyl-diphosphate synthase
C:Species: Bacillus stearothermophilus
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Jun-2000
C:Accession: JX0257
R:Koyama, T.; Odate, S.; Osabe, M.; Takeshita, A.; Yokoyama, K.; Uchida, M.; Nishino, T.
J. Biochem. 113, 335-363, 1993
A:Title: Thermotable farnesyl diphosphate synthase of Bacillus stearothermophilus: mole
A:Reference number: JX0257; MUID:93252758; PMID:8486607
A:Accession: JX0257
A:Molecule type: DNA
A:Residues: 1-297 <MOV>
A:Cross-references: GB:D13293; NID:G391609; PIDN:BA02551.1; PID:G391610
C:Genetics:
A:Start codon: GTG
C:Superfamily: Geranyltranstransferase
C:Keywords: transferase

Query Match      23.6%; Score 360.5; DB 2; Length 297;
Best Local Similarity 34.9%; Pred. No. 1.6e-22;
Matches 98; Conservative 41; Mismatches 129; Indels 13; Gaps 5;

QY 14 ISAEQLADIDSRLLDQLLPVQGER----DCVGAANREGTLAPGKIRPMLLALTARDLGC 69
DB 4 LSVQEPFLNQKQAVETALSRYIERLEGPAKLKAMAYSLEAGGKRIRPILLSTVRAUGK 53
QY 70 AISHGGLLDLACAVEMVHAASLIIDMPCKDDAQMERGRFTTHIQYGERHVAILAAVALLS 129
DB 64 DPVAG--LPVACAIEMHWYSLIHDDLPSMNDLLRRGKPTNKHVFGEMAILLAGDLIT 121
QY 130 KAFGVIAEA--EGLTPIATRAVSELSTAIGMOGLVQGGFKLSEGDKPSADAILTNQ 187
DB 122 YAFQLITEIDDERIPESVRILIERLAKAPEGVAGQADMEGEGKTLTISELEYIHR 181
QY 188 FKTSTLFCASTQWASIAANASCEARENLHRPSLDLGOAFOLLDDLT-D-----GMTDTGK 244
DB 182 HKTGRMLQVSVHAGALIGADARQTFELDEFAHLGLAFQIRDDILDIEGAEKIGKPVG 241
QY 245 QDA--GKSTLVNLGSGAVBERLRQHRLASEHLSAAQNG 283
DB 242 SDOSNNKATVPALLSLAGAKELKIAPIEAAQSHLRNADVDG 282

```

```

RESULT 14
F97685
geranyltransferase (AF203481) [imported] - Agrobacterium tumefaciens (strain C58, Cereon
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C/Accession: F97685
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markeiz, B.,
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359; MOID:21608551; PMID:11743194
A/Accession: F97685
A/Status: preliminary
A/Molecule type: DNA

```

```

>Residues: 1-304 <KUR>
>Cross-references: GB:AE007869; PIDN:AAK88439.1; PID:g15157936; GSPDB:GN00169
>Genetics:
>Gene: AGR_C_4933
>Map position: circular chromosome

Query Match      23.5%; Score 358.5; DB 2; Length 304;
Best Local Similarity 36.0%; Pred. No. 2.5e-22;
Matches 98; Conservative 44; Mismatches 111; Indels 19; Gaps 7;

2Y 17 ECLLADI---DSRLDQLLPVQGERDCVGAAMREGTLPAGKRIKIRPMLLLLTARDLGCASH 73
2b 20 EALLGHLISGEARADEITRPQNLE---AHEGVNLGKRLRPFLVIESVALLGGDAEA 75
2Y 74 GGLLDLACAVEMVHAASLILDDMPDQDDQMRGRPTHTQYGEHVAILAAVALLSKAFG 133
2b 76 G--LHVGAALCELCYSLVHDDLPAFMDDDLLRRGQPTVHRKFDDEATAILAGSLLTLAFD 133
2Y 134 VIAEAEGLTPIA---KTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSDAILLTNQPKT 190
2b 134 IIRADD--NPLAERKALVLSARAAGICGMAGGQALDLAAEKAPDEGIIITLQAMKT 191
2Y 191 STLFCASTQMASIAANASCEARENLRHFRSLDLGQAFOLLDDLTGQMTDT---GKDIHQDA 247
2b 192 GALLRFACEAGAILAGSNQAEORQLRFGEKIGLSFQIADDDLLDTADATMGKATGKDA 251
2Y 248 --GKSTLVNLLGSGAVEERLRFQHLRLASEHLS 277
2b 252 ARGKGTLLVALRGEAWAREKLOQVAAEASELLA 283

RESULT 15
AH2910
Geranyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH2910
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.;
  erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
  ; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoc, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
  ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2910
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAI43702.1; PID:g17741230; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ispA
A:Map position: circular chromosome
C:Superfamily: dimethylallyltransferase

Query Match      23.5%; Score 358.5; DB 2; Length 335;
Best Local Similarity 36.0%; Pred. No. 2.8e-22;
Matches 98; Conservative 44; Mismatches 111; Indels 19; Gaps 7;

QY 17 ECLLADI---DSRLDQLLPVQGERDCVGAAMREGTLPAGKRIKIRPMLLLLTARDLGCASH 73
2b 51 EALLGHLISGEARADEITRPQNLE---AHEGVNLGKRLRPFLVIESVALLGGDAEA 106
QY 74 GGLLDLACAVEMVHAASLILDDMPDQDDQMRGRPTHTQYGEHVAILAAVALLSKAFG 133
2b 107 G--LHVGAALCELCYSLVHDDLPAFMDDDLLRRGQPTVHRKFDDEATAILAGSLLTLAFD 164
QY 134 VIAEAEGLTPIA---KTRAVSELSTAIGMQLVQGFQKDLSEGDKPRSDAILLTNQPKT 190
2b 165 IIRADD--NPLAERKALVLSARAAGICGMAGGQALDLAAEKAPDEGIIITLQAMKT 222
QY 191 STLFCASTQMASIAANASCEARENLRHFRSLDLGQAFOLLDDLTGQMTDT---GKDIHQDA 247
```

```

223 GALLRFACEAGAILAGSNQAEORQLRFGEKIGLSFQIADDDLLDTADATMGKATGKDA 282
248 --GKSTLVNLLGSGAVEERLRFQHLRLASEHLS 277
283 ARGKGTLLVALRGEAWAREKLOQVAAEASELLA 314

Search completed: February 29, 2004, 14:52:42
Job time : 9.43386 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

% protein - protein search, using sw model

un on: February 29, 2004, 14:27:18 ; Search time 5.10674 Seconds

(without alignments)
3089.496 Million cell updates/sec

itle: US-09-941-947A-26

erfect score: 1526

sequence: 1 LTVCAKKHVHTGISABQL.....HSTTQLFIQAWPDKKLAWS 303

coring table: BLQSUM62

Gapop 10.0 , Gapext 0.5

sarched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1344.5	88.1	302	1	CRTE_PANAN
2	762.5	50.0	307	1	CRTE_ERWHE
3	372	24.4	369	1	GGPP_CAPAN
4	369	24.2	371	1	GGPP_ARATH
5	364.5	23.9	366	1	GGPP_SINAL
6	360.5	23.6	297	1	ISPA_BACST
7	346.5	22.7	272	1	ISPA_BACSU
8	345.5	22.6	357	1	GGPP_CATRO
9	332	21.8	289	1	ISPA_SCOLI
10	320	21.0	332	1	ISPA_RHISN
11	316	20.7	332	1	ISPA_BRAJA
12	313.5	20.5	300	1	CRTE_CVAPA
13	309.5	20.3	295	1	ISPA_HAEIN
14	307	20.1	288	1	CRTE_RHOSH
15	304.5	20.0	291	1	ISPA_MICLU
16	300	19.7	294	1	ISPA_BUCAP
17	297.5	19.5	289	1	CRTE_RHOCA
18	277	18.2	282	1	ISPA_BUCAI
19	272	17.8	329	1	ISPB_HAEIN
20	249.5	16.3	262	1	ISPA_AQUAS
21	246	16.1	334	1	ISDA_METTM
22	244	16.0	325	1	ISDA_METTH
23	243.5	16.0	323	1	PREA_SVNY3
24	242	15.9	323	1	ISPB_ECOLI
25	234	15.3	320	1	HEP2_BACST
26	233.5	15.3	332	1	GGPP_SULSO
27	232	15.2	323	1	PREA_CVAPA
28	231	15.1	359	1	GGPP_WCTCT
29	230	15.1	327	1	ISDA_MTLAJ
30	219.5	14.4	330	1	GGPP_MTLAC
31	203.5	13.3	348	1	HEP2_BACSU
32	194.5	12.7	378	1	DPS_SCHPO
33	193.5	12.7	323	1	PREA_CVAPA

ALIGNMENTS

RESULT 1

```

CRTE_PANAN
ID CRTE_PANAN STANDARD; PRT; 302 AA.
AC P21684;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP
DE synthetase) (Farnesyltransferase).
GN CRTE.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=353;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20D3;
RX MEDLINE=91072214; PubMed=2254247;
RA Nakamura K., Harashima K.;
RA Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,
RT "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway
RT by functional analysis of gene products expressed in Escherichia
RT coli.";
RL J. Bacteriol. 172:6704-6712(1990).
CC -1- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -1- PATHWAY: Carotenoid biosynthesis.
CC -1- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; D90087; BAAL4124.1;
CC PIR; A37802; A37802.
CC InterPro; IPR000092; Polyrenyl_synt.
CC InterPro; IPR008949; Terpenoid synth.
CC Pfam; PF00348; polyrenyl_synt; 1.
CC PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
CC PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
CC Carotenoid biosynthesis; Isoprene biosynthesis; Transferase.
CC SEQUENCE 302 AA; 32563 MW; CADB04695D2EAA4A CRC64;

```

Query Match 88.1%; Score 1344.5; DB 1; Length 302;
Best Local Similarity 88.4%; Pred. No. 2.6e-101;
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

```

QY 1 LTVCAKKHVHTGISABQLIDSLDQLPVGGRDCVGAARECTAPGKRIRPML 60
Db 1 MFTVCAKKHVHTGADAAEQQLADIDRRDQLLPVEGRDVGGAAREGALAPGKRIRPML 60
QY 61 LITARDLGAISHGGLLDLACAVEMVHAASLLIDDMPCMDDAQMRGRPTTIHTQYGEHVA 120

```


Db 61 ULTARDLGCANSHGGLLDLACAVEMVHAASLIIDMPCMDAKLRGRPTIHSYGRHVA 120
 QY 121 ILAAVALLSKAFGVIAEAGTPIAKTRAVSELSTAIGMQLVQOGPKOLSEGDKPRSD 180
 Db 121 ILAAVALLSKAFGVIAEAGTPIAKTRAVSELSTAIGMQLVQOGPKOLSEGDKPRSD 180
 QY 181 AILLTNQFKTSTLFCASQOMASIAANASCEARENLRHFSLDLQOAFOLLDDLTGMDTDC 240
 Db 181 AILLTNQFKTSTLFCASQOMASIAANASCEARENLRHFSLDLQOAFOLLDDLTGMDTDC 240
 QY 241 KDINQDAGKSTLVNLLGSGAVSERLRQHLRLASHLASAACONGHSTTQLFIQAWFDKCLA 300
 Db 241 KDSNODAGKSTLVNLLGSGAVSERLRQHLRLASHLASAACONGHSTTQLFIQAWFDKCLA 299
 QY 301 AVS 303
 Db 300 AVS 302

RESULT 2

TEETE ERWHE
 ID CRUE ERWHE STANDARD; PRT; 307 AA.
 AC P22873;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP synthetase) (Farnesyltransferase).
 EN CRTE.
 OS Erwinia herbicola.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=549;
 RN [1] SEQUENCE FROM N.A.
 RC STRAIN=EH010;
 RX MEDLINE=91088634; PubMed=2263648;
 RA Armstrong G.A., Albert M., Hearst J.E.;
 RT "Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).
 RN [2]
 RP FUNCTION
 RX MEDLINE=92357711; PubMed=1495965;
 RA Math S.K., Hearst J.E., Poulter C.D.;
 RT "The crtB gene in Erwinia herbicola encodes geranylgeranyl diphosphate synthase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6761-6764(1992).
 CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + Geranylgeranyl diphosphate.
 CC -!- PATHWAY: Carotenoid biosynthesis.
 CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M38424; AAA24819.1; -;
 CC EMBL; M87280; AAA64977.1; -;
 CC PIR; C39273;
 CC InterPro; IPRO00092; Polyrenyl_synth.
 CC InterPro; IPRO08949; Terpenoid_synth.
 CC Pfam; PF0348; polyrenyl_synth; 1.
 CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
 CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
 CC Carotenoid biosynthesis; Isoprene biosynthesis; Transferase.
 CC SEQUENCE 307 AA; 33242 MW; 6A534C6194CE9F59 CRC64;

Query Match 50.0%; Score 762.5; DB 1; Length 307;
 Best Local Similarity 56.0%; Pred. No. 2.4e-54;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
 QY 17 EQLADIDSRDLQQLPVQGERDCVGAARREGTLAPGRIRPIMILLTARDIGCAISHGGL 76
 Db 16 EVMRQSIDDLHAGLLPETDSQDIVSLARREGVMAFGKIRPULMLAARDLRYQGSMTL 75
 QY 77 LDLACAVEMVHAASLIIDMPCMDDAQVRRGRPTIHYGEHVAILAAVALLSKAFGVIA 136
 Db 76 LDLACAVELTHTASLMDLMDPCMDNRELRRGQPTTHKKFGESVAILASVGLLSKAFGLIA 135
 QY 137 EAEGLTPIAKTRAVSELSTAIGMQLVQOGPKOLSEGDKPRSDAILTNQFKTSTLPCA 196
 Db 136 ATGDLPGERRAQAVNELSTAVGVLGQFRLDNDAAALDRTPDAILSTNHLKTGLFSA 195
 QY 197 STOMASTAANASCEARENLRHFSLDLQOAFOLLDDLTGMDTDCINQDAGKSTLVNLL 256
 Db 196 MLQIVAIASASSPSTRETTLAFALDPQAFQLLDLDRDDHPETGKDRNKDAGKSTLVNRL 255
 QY 257 GSGAVEERLRQHLRLASHLASAACONGHSTTQLFIQAWFDKCLA 300
 Db 256 GADAAQKLRHIDSADKHLTFACPGGAIHQ.PWHLWFGHLLA 298
 RESULT 3
 ID GGPP CAPAN STANDARD; PRT; 369 AA.
 AC P80042;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
 EN GGPS1.
 OS Capsicum annuum (Bell pepper).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 CC Lamiales; Solanales; Solanaceae; Capsicum.
 OX NCBI_TaxID=4072;
 RN [1] SEQUENCE FROM N.A.
 RC STRAIN=cv. Lamuyo; TISSUE=Fruit;
 RX MEDLINE=93272043; PubMed=1303794;
 RA Kuntz M., Roemer S., Suire C., Huguency P., Weil J.H., Schantz R., Camara B.;
 RT "Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme activity and transcript level during fruit ripening.";
 RL Plant J. 2:25-34(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yolo Wonder;
 RX MEDLINE=95195169; PubMed=7888631;
 RA Badillo A., Steppuhn J., Deruere J., Camara B., Kuntz M.;
 RT "Structure of a functional geranylgeranyl pyrophosphate synthase gene from Capsicum annuum.";
 RL Plant Mol. Biol. 27:425-428(1995).
 CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.
 CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans-trans-farnesyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.
 CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key enzyme in plant terpenoid biosynthesis.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

```

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; X80267; CAA56554.1; --
PIR; S53722; S53722; Polyprenyl synth.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00348; polyprenyl synth; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
Transferase; Isoprene biosynthesis; Carotenoid biosynthesis;
Chloroplast; Transit peptide. CHLOROPLAST.
CHAIN 1 369 GERANYLGERANYL PYROPHOSPHATE SYNTHETASE.
SEQUENCE 369 AA; 40173 MW; 2D527F5B43A29C6F CRC64;
Query Match 24.4%; Score 372; DB 1; Length 369;
Best Local Similarity 34.5%; Pred. No. 9.6e-23;
Matches 99; Conservative 52; Mismatches 108; Indels 28; Gaps 7;
/ 23 IDRLDQLPVGGRDCVGAAMREGLTAPGRIRPMLLLTARDLGCAISHGGLDLACA 82
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: 89 VNKADELAIIV-EPHVIHEAMRYSLAGGRVREMLCLAAACELVGG--NQENAVAACA 145
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 83 VEMVRAASLIDMPCDDAOWERGPRTHQYGEHVAILAVALLSKAF-GVTAAREGL 141
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: 146 VENHTMSLHDDPCMDNDLRRGKPTNKHLYGDEVAVLGSDLSLAFAPHHVNSTAGV 205
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 142 TPIAKTRAVSELSTAIGMGLVGGQFQDLX-EGDKPRSAAILITNPQKTLFCASTOM 200
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: 206 TFSRVGAVAEALAKSIGTEGLVAGQVADIKCTGNASVLETLFHFVHKTAALLESSVWL 265
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 201 ASIANASCEARENLRSLDGOAFQQLDLDTGMDTGTGDIQWDACKSTLVN----- 254
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: 266 GAILGGTGVVEVKLRARRFARCIIGLFWQVDILD-VTKSELKTAGKAGLDVVDKTTYPK 324
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
/ 255 LKSGAVERLRQHLRLASEHLSAACQNGHSTTQLFQAFQKGLAA 301
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: 325 LKGLKAKFAELNREAKQLEG-----FDSRKAA 355
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 4
PFP ARATH STANDARD; PRT; 371 AA.
P34B02; 023201;
01-FEB-1994 (Rel. 28, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP
synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1);
Geranyltransferase (EC 2.5.1.10); Farnesyltransferase
(EC 2.5.1.29)]
GGPS1 OR AT4G36810.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI TaxID=3702;
[1]
SEQUENCE FROM N.A.
MEDLINE=94286754; PubMed=8016276;
Scolnik P.A., Bartley G.E.;
"Nucleotide sequence of an Arabidopsis cDNA for geranylgeranyl
pyrophosphate synthase."
Plant Physiol. 104:1469-1470(1994).
[2]
SEQUENCE FROM N.A.

```

```

RC STRAIN=cv, Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Scliekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansgore W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen J., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Heiseel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.A., McCullagh B., Bilham L., Robben J.,
RA van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Bernais S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacón D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Meves H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Farnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes K., Abu-Threideh J.,
RA Stonking T., Kallick J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spieth J., Ryan B., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanich M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana."
RL Nature 402:769-777(1999).
CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of
CC Ipp onto DMAPP to form geranylgeranyl pyrophosphate.
CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
CC diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key
CC enzyme in plant terpenoid biosynthesis.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; L25813; AAA32797.1; --
CC EMBL; Z99708; CAB16803.1; --
CC EMBL; AL161590; CAB80347.1; --

```


X MEDLINE=95001990; PubMed=7918490;
A Koyana T., Obata S., Saito K., Takeshita-Koike A., Ogura K.;
T "Structural and functional roles of the cysteine residues of Bacillus
I steatotherophilus farnesyl diphosphate synthase.";
L Biochemistry 33:12644-12648(1994).
C C-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
C = diphosphate + trans,trans-farnesyl diphosphate.
C -!- SUBCELLULAR LOCATION: Cytoplasmic.
C -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
R EMBL: D13293; BAA02551.1; -
R ENBL: S72629; AAB32272.1; -
R ENBL: S72630; AAB32273.2; ALT_SEQ.
R EMBL: S72633; AAB32274.1; -
R EMBL: S72635; AAB32275.2; ALT_SEQ.
R PIR: JX0257; JX0257.
R InterPro: IPR000092; Polyprenyl_synth.
R InterPro: IPR008949; Terpenoid_synth.
R Pfam: PF00348; polyprenyl_synth; 1.
R PROSITE: PS00444; POLYPRENYL SYNTHET 2; 1.
R PROSITE: PS00723; POLYPRENYL SYNTHET 1; 1.
W Transferase; Isoprene biosynthesis.
T MUTAGEN 73 73 C->F.S: NO LOSS OF ACTIVITY.
T MUTAGEN 289 289 C->F.S: NO LOSS OF ACTIVITY.
T SEQUENCE 297 AA; 32310 MW; 0F921C3F029EBB6 CRC64;
Q
Query Match 23.6%; Score 360.5; DB 1; Length 297;
Best Local Similarity 34.9%; Pred. No. 6.3e-22;
Matches 99; Conservative 41; Mismatches 129; Indels 13; Gaps 5;
Y 14 ISAPOLLADISRLDQLLPQGER-----DCVGNAREGLTAPGKIRPMMLLTARDGCG 69
b 4 LSVQEFNEQKQAVETALSGRYLERLEGPAKIKKAMAYSLEAGGKIRPULLUSTVRLGK 63
Y 70 AISHGGILDLACAVEMVHAASLILDMPCVDDAQMRGERPTIHTQYGEHVAILAAVALLS 129
b 64 DPAVG--LPVACAIEIMHYSLIHDDLPSVNDLLRGKPTNKHVFGEMAILAGDGLT 121
Y 130 KAFGVIAEA--EGLTPIAKTRAVSELSTALGMGLVQGFOLSLSGDKPRSDAILLTNQ 187
b 122 YAFQLITEIDDERIPPSVRLRIETRIAKAAGPEGMVAGQAADMEGEGKTLTISELEYIHR 181.
Y 188 FKTSTLFCASTOWASTAANASCEARENLRFLSLDGAQFOLLDDLT--GMDTGTGKDN 244
b 182 HKTGMQLQSVHAGALIGGDARQTELEFPAHLGLAQIRDDILDIEGAEEKIGKPVG 241
Y 245 QDA--GKSTLWNLGSGAVEERLRQHLRLASEHLSAACQNG 283
b 242 SDQSNKATVPALLSLAGAKEKLAFLHIAEQRLRNADVGG 282
RESULT 7
SPA BACSU STANDARD; PRT; 272 AA.
D IGPA BACSU
C P54383;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
E (FPP synthase).
N YQID OR BSU24280.
S Bacillus subtilis.
C Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
X NCBI_TaxID=1423;
N [1]

RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno K., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogatawara N., Mosser J., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourschi L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly B.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan B., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrero P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tateso V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Mambatt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
Nature 390:249-256(1997).
-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
= diphosphate + trans,trans-farnesyl diphosphate.
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
R EMBL: D84432; BAA12575.1; -
R EMBL: Z99116; CAB14359.1; -
R PIR: A69961; A69961.
R Subtilist; BG1714; YQID.
R InterPro: IPR000092; Polyprenyl_synth.
R InterPro: IPR008949; Terpenoid_synth.
R Pfam: PF00348; polyprenyl_synth; 1.
R PROSITE: PS00444; POLYPRENYL SYNTHET 2; 1.
R PROSITE: PS00723; POLYPRENYL SYNTHET 1; 1.
W Transferase; Isoprene biosynthesis; Complete proteome.
K SEQUENCE 272 AA; 29740 MW; 55A26E9A9895D67C CRC64;
S
Query Match 22.7%; Score 346.5; DB 1; Length 272;
Best Local Similarity 34.0%; Pred. No. 7.6e-21;
Matches 91; Conservative 52; Mismatches 108; Indels 17; Gaps 7;
N

20 LADIDSLDQLLPVQGER----DCVGAAMEGTLAPCKRIRPMLLLLTARDLGCALSHGG 75
 9 LADRKKTENGLSVYTEKLDMPDPSLKKSMLYSLQAGGKRLRPLVLAVLNAYKSEKDG- 67
 76 LLDLACAVETHAASLILDDPCNDQMRGRPTTHQYGEHVAILAAVALLSKAFGVI 135
 68 -IPGCAVEMHTYLSLHDDDFCMMDDDLRGKPTNKHVGEATAVLAGDELLESFKLI 126
 136 AE--AEGTLPIAKTRAVSELSAIGMGLVQGFQKDLSEGDKPSADAILLTNQFKTSL 193
 127 TSHVSDVSAEKRLRLYNELISAAGTEGMVGQVADMEAGNRQVTLLEESIHAKTSKL 186
 194 --FCASTOMASIAANASCEARENLRSLDLGQAFQLLDLTD---GMTDTGDIQND- 247
 187 LGFCVIA--GAILADAPEDTETRTSSHIGIGFQIRDDILDLEGSEKIGKRVGSDTT 244
 248 -GKSTLVNLLSGAVEERLRQHLRIASE 274
 245 NDKSTYPSLLSLEGAKKHLDVHKEAND 272

RESULT 8

3GPP CATRO STANDARD; PRT; 357 AA.
 ID -GGPP CATRO STANDARD; PRT; 357 AA.
 AC Q42698;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
 DE GGPS1 OR GC1.
 DE Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 DE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 DE Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 DE lamids; Gentianales; Apocynaceae; Rauvolfioideae; Vincet; Catharanthus.
 DE NCBI TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. G. Don c20;
 RA Bantignies B., Liboz T., Ambid C.;
 RT "Nucleotide sequence of a Catharanthus roseus geranylgeranyl pyrophosphate synthase gene."
 RL (in) Plant Gene Register PGR95-119.
 CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.
 CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Trans,trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.
 CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key enzyme in plant terpenoid biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X92893; CAA63486.1; --
 PIR; T09966; T09966
 InterPro; IPR000092; Polyrenyl_synth.
 DR Pfam; PF00348; polyrenyl_synth; 1.
 DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.

DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 KW Transferase; Isoprene biosynthesis; Carotenoid biosynthesis;
 KW Chloroplast; Transf peptide.
 FT TRANSIT 1 40 CHLOROPLAST (POTENTIAL).
 FT CHAIN 41 357 GERANYLGERANYL PYROPHOSPHATE SYNTHETASE.
 SQ SEQUENCE 357 AA; 38786 MW; 81C52FDEA1E06FA8 CRC64;
 Query Match 22.6%; Score 345.5; DB 1; Length 357;
 Best Local Similarity 37.9%; Pred. No. 1.3e-200; Indels 11; Gaps 5;
 Matches 92; Conservative 40; Mismatches 100;
 QY 43 AMRSGTLPAGKRIRPMLLLLTARDLGCALSHGGILDLACAVEMVHAASLILDDPCMDDA 102
 DB 96 SMRYSLLAGGKRVKRLCIAACELFGGTESVA--MPSACAVEMHTMTSLMEDDLPCMDND 153
 QY 103 QMRGRPTTHQYGEHVAILAAVALLSKAFGVTABA-EGTLPIAKTRAVSELSAIGMGL 161
 DB 154 DLRRGKPTNKHVGEATAVLAGDELLESFKLIATATKGVSSERIVRVGELAKCIGSEG 213
 QY 162 LVQGFQKDL-SEGDKPSADAILLTNQFKTSLFCASTOMASIAANASCEARENLRHFSL 220
 DB 214 LVAGQVYDVCSEGLADVGLHLEFIHKTAALEGSVVLGAVGGANDQISKLRKFAR 273
 QY 221 DLGQAFQLLDLTDGMTDTGKDIQNDAGKS-----TLVNLLSGAVEERLRQHLRIASE 274
 DB 274 CIGLLFQVDDILD-VTKSSQELGKTAGKDLVADKVTYPKLLGIDKRSREFAEKLNEAQE 332
 QY 275 HLS 277
 DB 333 QLA 335
 RESULT 9
 ID ISPA_ECOLI STANDARD; PRT; 299 AA.
 AC P22939;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
 DE (FPP synthase).
 DE ISPA OR B0421.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=9210228; PubMed=2089044;
 RA Fujisaki S., Hara H., Nishimura Y., Horiuchi K., Nishino T.;
 RT "Cloning and nucleotide sequence of the ispa gene responsible for farnesyl diphosphate synthase activity in Escherichia coli."
 RL J. Biochem. 108:995-1000(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Duncan M., Federspiel M., Hyman R., Kaiman S., Komp C., Kurdi O., Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

-!- SIMILARITY: Belongs to the FPP/GPPP synthetase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; D00694; BAA00599.1; -;
 EMBL; AE000148; AAC73524.1; -;
 EMBL; U82664; AAB40177.1; -;
 PIR; JQ0665; JQ0665.
 SWISS-2DPAGE; P22939; COLI.
 EcoGene; EG10508; ispa.
 InterPro; IPR000092; Polyprenyl_synth.
 InterPro; IPR008949; Terpenoid_synth.
 Pfam; PF00348; polyprenyl_synth.1.
 PROSITE; PS00444; POLYPRENYL_SYNTHET_1; 1.
 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 TRANSFERASE; Isoprene biosynthesis; Complete proteome.
 SEQUENCE 299 AA; 32159 MW; 15BAD05E135060CA CRC64;
 Query Match 21.8%; Score 332; DB 1; Length 299;
 Best Local Similarity 33.8%; Pred. No. 1.3e-19;
 Matches 101; Conservative 46; Mismatches 124; Indels 28; Gaps 8;
 1 LTVCAKHEVHLTGISAEQLADISRLDQLLPVOCERDCVGAAMREGTLAPGKIRPMLL 60
 7 LEACVKQ-----ANGAL-----SRFLAPLFQNTP--VVEWYQYGLLGGKRLRFLV 52
 61 LITARDLCAISHGGLLDLACAVENMVAASLIIDMPCQDDQAMRGRPTIHTQYGEHVA 120
 53 YATGHMFG--VSTNTLDAPAAVECTHAYSLIHDLPAMDDDLRLRGLPTCHVKFGEANA 110
 121 ILAAVALLSKAFGVIAEGLTPIAKTR---AVSELSTAIGMGLVQGFQDLSEGDKP 176
 111 ILAGDALQTLAPSLISDAD--MPEVSDRDRISMISELASASGIAGMCGGQALDLDAEGKH 168
 177 RSADAILLTNQFTSLTFCASTQMASIAA--NASCEARENLRHPSLDLQAPQLDDLTG 235
 169 VPLDALERHHRKHTGALLRAAVRIGALSAGDKRRALPVLQKYAESIGLAFQVQEDILDV 228
 236 MDTDTG----XDINDACKSLVNLGSGAVEERLRQLRLRLASEHLASAQNGHSTTQL 289
 229 VEDTATLKGKQAGDQQLGKSTYPAALLGLRQARKARDLIDDRSLKQLAEQSLDTSAL 287
 RESULT 10
 ID ISPA RHISN STANDARD; PRT; 332 AA.
 AC Q45220;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable geranyltranstransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase) (FPP synthase).
 GN Y4KU.
 OS Rhizobium sp. (strain NGR234).
 PE Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 NCBI_TaxID=394;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=97305956; PubMed=9163424;
 Freiburg C.A., Feilay R., Bairoch A., Broughton W.J., Rosenthal A., Perret X.;
 "Molecular basis of symbiosis between Rhizobium and legumes."
 Nature 387:394-401 (1997).
 -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.

-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 -!- SIMILARITY: Belongs to the FPP/GPPP synthetase family.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; AE000082; AAB91752.1; -;
 PIR; T10875; T10875.
 InterPro; IPR000092; Polyprenyl_synth.
 InterPro; IPR008949; Terpenoid_synth.
 Pfam; PF00348; polyprenyl_synth.1.
 PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
 PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
 TRANSFERASE; Isoprene biosynthesis; Plasmid.
 SEQUENCE 332 AA; 34688 MW; EB68C1547449AA87 CRC64;
 Query Match 21.0%; Score 320; DB 1; Length 332;
 Best Local Similarity 35.4%; Pred. No. 1.3e-18;
 Matches 102; Conservative 40; Mismatches 106; Indels 40; Gaps 10;
 23 IDSRLDQLLPVQ--GERDCVGAAMREGTLAPGKIRPMLLTLTARDLCAISHGGLLDLA 80
 44 VEQALRLLCADHEDGTELM-AMRYATLHGKGRTR-ALLCLAGALADTFAH-ILDDVG 100
 81 CAVENMVAASLIIDMPCQDDQAMRGRPTIHTQYGEHVAIILAAVALLSKAFGVIA--EA 138
 101 ALEEMWHAETLVHDDLPAWDDVLRGLPTVHVKEGENTAILVGDALQAHAFLLTSLASLDA 160
 139 EGLTPIAKTRAVSELSTAIGMGLVQGFQDLSEGDKPRSADAILLTNQFTSLTFCAST 198
 161 PDENRIALVR---ELAAVSAEAGAGGQAMDLSLVGKVELDRIVAMRMCKGALVRASV 217
 199 QWASIAAANASCEARENLRH---RPSLDLQAPQLDDLTGKWTDT---GKINDODAG--K 249
 218 RMGALCAEAEDADATLYCALDHYSCFGLALQVDDILDATATLGTGTPSKDAAQK 277
 250 STLNVLLSGAVE-----ERLRQLRLRLASEHL 276
 278 PTCASIMGLQAARQFALDLCAGEATAIPLGPRAEERLAQMLQASAVL 325
 RESULT 11
 ID ISPA BRAJA STANDARD; PRT; 332 AA.
 AC Q45220;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable geranyltranstransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase) (FPP synthase).
 GN FPPS OR BLR2148.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 NCBI_TaxID=375;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=USDA 110;
 Tully R.E., Keister D.L.;
 "Cloning and mutagenesis of a cytochrome P-450 locus from Bradyrhizobium japonicum that is expressed anaerobically and symbiotically."
 Appl. Environ. Microbiol. 59:4136-4142 (1993).
 [2]
 SEQUENCE FROM N.A.
 STRAIN=USDA 110;
 MEDLINE=98322110; PubMed=9655913;
 Tully R.E., van Berkum P., Lovins K.W., Keister D.L.;

RT "Identification and sequencing of a cytochrome P450 gene cluster from
 RT *Bradyrhizobium japonicum*".
 RL Biochim. Biophys. Acta 1396:243-255 (1998).

RW (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 UA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 UA Sasamoto S., Watanabe A., Idegawa K., Iriiguchi M., Kawashima K.,
 UA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 UA Tabata S.;
 XT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 XT *Bradyrhizobium japonicum* USDA110.";
 XL DNA Res. 9:189-197(2002).
 XC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
 XC = diphosphate + trans-trans-farnesyl diphosphate.
 XC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 XC -!- SIMILARITY: Belongs to the PPP/GGPP synthetase family.
 XC
 XC This SWISS-PROT entry is copyright. It is produced through a collaboration
 XC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 XC the European Bioinformatics Institute. There are no restrictions on its
 XC use by non-profit institutions as long as its content is in no way
 XC modified and this statement is not removed. Usage by and for commercial
 XC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 XC or send an email to license@sib-sib.ch).
 XC
 XR EMBL; U12678; AAC28894.1;
 XR EMBL; AP005942; BAC47413.1; ALT_INIT.
 XR PIR; I40213; I40213.
 XR InterPro; IPR000092; Polyrenyl_synt.
 XR InterPro; IPR008949; Terpenoid_synt.
 XR Pfam; PF00348; Polyrenyl_synt; 1.
 XR PROSITE; PS00444; POLYPRENYL_SYNTHET 2; 1.
 XR PROSITE; PS00723; POLYPRENYL_SYNTHET 1; 1.
 RW Transferase; Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 332 AA; 34638 MW; 47644FAF0220CE4C CRC64;

Query Match 20.7%; Score 316; DB 1; Length 332;
 Best Local Similarity 33.4%; Pred. No. 2.9e-18;
 Matches 107; Conservative 45; Mismatches 128; Indels 40; Gaps 11;
 2Y 8 HVHLTGTSABQLLA-----DIDSRLDQLLPVQ--GERDVCGAAM 44
 2E 8 HDDRAGVSANGILQAQAGSGLLPPIWQDGAKRVEEARILCAEDDGETELM-AAM 66
 2Y 45 REGTAPGKRIRPMLLLTARDLCAISHGGLDLACAVENVHAASLIIDMPCMDAQM 104
 DB 67 RYATLHGKRTK-ALLCLAAAGALADTPAH-MLDDVGAALIMHHAETLVHDDLPAMD 124
 2Y 105 RRGRTHTQYGERHVAALAAVALLSKAPGVIA--EAEGLTPIAKTRAVSELSTAIGMQL 162
 DB 125 RRGLTPTVHKVGEATILVGDALQAHPLTLASINAFCDSPIALVR---ELAQVSAEGA 181
 2Y 163 VQGQFKDLSGDKPRSDAILLTNQKSTLFCASTOMASIAA---NASCEARE-NLHRP 218
 DB 182 AGGQAIDLSLVGKGVHEDRIIVAMHRMKGALVRASVRMGALCAVGVAHAALYCALDHY 241
 2Y 219 SLDLQAFQLDDLTGMDTDT--GKDINODAG--KSTLVNLLGSGAVEERLQHLRLAS 273
 DB 242 SACSGLAQVDDILDVTAUTAAAGKTPGKDNAAQKPTCASIMGLQBAROFALDILLRDAG 301
 2Y 274 EHLSAACQNGHSTTQLPFOA 293
 DB 302 EALAPLGPRAERLAQLIORA 321

RESULT 12
 CITE CYAPA
 ID CITE CYAPA STANDARD; PRT; 300 AA.
 AC P48368;
 DT 01-FEB-1996 (Rel. 33; Created)
 DT 01-FEB-1996 (Rel. 33; Last sequence update)

DT 10-OCT-2003 (Rel. 42; Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase homolog (EC 2.5.1.29) (GGPP
 DE synthetase) [Farnesyltransferase].
 GN CRIE.
 OS Cyanophora paradoxa.
 OC Cyanelle.
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
 OX NCBI_TaxID=2762;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
 RA Bryant D.A.;
 RT "Nucleotide sequence of the cyanelle DNA from *Cyanophora paradoxa*."
 RL Plant Mol. Biol. Rep. 13:327-332(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
 RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
 RT "The complete sequence of the cyanelle genome of *Cyanophora paradoxa*:
 RT the genetic complexity of a primitive plastid."
 RL (In) Schenk H.B.A., Herrmann R., Jeon K.W., Mueller N.E.,
 RL Schwemmler W. (eds.);
 RL Bukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
 RL (1997).
 CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranylgeranyl diphosphate.
 CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 CC -!- SIMILARITY: Belongs to the PPP/GGPP synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; U30821; AAA81312.1; --
 DR PIR; T06969; T06969.
 DR InterPro; IPR000092; Polyrenyl_synt.
 DR InterPro; IPR008949; Terpenoid_synt.
 DR Pfam; PF00348; Polyrenyl_synt; 1.
 DR PROSITE; PS00444; POLYPRENYL_SYNTHET 2; 1.
 DR PROSITE; PS00723; POLYPRENYL_SYNTHET 1; 1.
 RW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 KW Isoprene biosynthesis; Transferase; Cyanelle.
 SQ SEQUENCE 300 AA; 33008 MW; 05950D0E3BA04934 CRC64;
 Query Match 20.5%; Score 313.5; DB 1; Length 300;
 Best Local Similarity 33.2%; Pred. No. 3.9e-18;
 Matches 87; Conservative 54; Mismatches 102; Indels 19; Gaps 8;
 2Y 23 IDSRLDQLLPVQGRDVCVGAAMREGTAPGKRIRPMLLLTARDLCAISHGGL---LDL 79
 DB 21 VEDTLNKSIP-RGNPTFIYDSIRYSLSAGGKRIRPILCLAS-----CELAGGTMEVALPT 74
 2Y 80 ACAYENVHAASLIIDMPCMDAQMREGRTHTQYGERHVAALAAVALLSKAPGVIA-EA 138
 DB 75 ACALMTHTLSLIHDDLPAMDNDSYRKGKPTNHIIVGEDLAILAGDALLAYAFEFIATQT 134
 2Y 139 EGLTPIAKTRAVSELSTAIGMQLVQGFQKDL-SEGDKPRSDAILLTNQKSTLFCAS 197
 DB 135 KNVPADLTIVKIVQAVSHSVTTSGLVGGQIIDLSESGKSDTTLTFLNFIHTKTGALLEAA 194
 2Y 198 TQMASIANASCEARENLHRSPLDLGOAFQLDDLTGMDT---DTGKDINOD--AGKSTL 252
 DB 195 VLSGALLAGKEKDMNRFYRQAQNIQAFQIIDDVLDIISTEEKLSIGKDLKTQKATY 254
 2Y 253 VNLGSGAVEERLQHLRLASE 274

255 PSFVG---VERSIOKQAEELVEE 273

RESULT 13

ISPA_HAEIN STANDARD; PRT; 295 AA.
P45204;
01-NOV-1995 (Rel. 32, Created)
01-NOV-1995 (Rel. 32, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Geranyltranstransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
(ZPP synthase).
159A OR H1438;
Haemophilus influenzae.
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
STRAIN-Rd / KW20 / ATCC 51807;
MEDLINE=95350630; PubMed=7542800;
Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
Venter J.C.;
"Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd";
Science 269:496-512(1995).

-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
= diphosphate + trans,trans-farnesyl diphosphate.
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U02822; AAC23087.1; --
PIR; C64123; C64123.

TIGR; H1438; --
InterPro; IPR000092; Polyprenyl_synth.
InterPro; IPR008949; Terpenoid synth.

Pfam; PF00348; polyprenyl synth; 1.

PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.

PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.

W TRANSFERASE; Isoprene biosynthesis; Complete proteome.

Q SEQUENCE 295 AA; 32407 MW; A64DB50266B670F CRC64;

Query Match 20.3%; Score 309.5; DB 1; Length 295;

Best Local Similarity 34.1%; Pred. No. 8.1e-18;

Matches 88; Conservative 46; Mismatches 107; Indels 17; Gaps 7;

20 LADISRLDQLLPVQGE-----RDCVGAMRECTLAPGKRIRPMLLLLTARDLQCAISHG 74

8 LQOVQTRINRFLEAQFEGIESHNAPLLEAMKAYALLGGKVRPFLWATGQMLG---AEK 64

75 GLLD-LACAVEMVHAASLTIDMPQMDQMRGRPTHTQVGHVAITAAVALLSKAFG 133

65 QTLDYAAAIAEIAHVSLLHDLPAWDDNLERRGHTCHIQFDEATALLAGDALQSFAFE 124

134 VIAEAGLLPIAKTAVSELSTAIGXQGLVQVQFNDLSEGDKPSADAILTNQFKSTL 193

125 ILTKPTNISTEOKLALIQLAGAGVQGMCLGQSLSLSEHKOISLSELELHNKTKGAL 184

OY 194 FCASTOMASIANASCEAR--ENLHRFSLDGLQAFQLDDLT---GMTTGKDI--NQD 246
DB 185 LIAALKGLFICSPHFTDKRLQSLTQYAEIAGLAFQVQDDILDIEGDSABETGKVGADLD 244

OY 247 ACKSTLVNLG-SGAVER 263

DB 245 LDKSTYPKLLGLSGAKQK 262

RESULT 14

CRTE RHOSH STANDARD; PRT; 288 AA.
AC P54976; Q9RFC5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP
synthetase) (farnesyltranstransferase).
GN CRTE.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=95338278; PubMed=7721699;
Lang H.F., Cogdell R.J., Takaichi S., Hunter C.N.;
"Complete DNA sequence, specific TMS insertion map, and gene
assignment of the carotenoid biosynthesis pathway of Rhodobacter
RT sphaeroides";
RT J. Bacteriol. 177:2064-2073(1995).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=20115911; PubMed=10648776;
Choudhary M., Kaplan S.;
"DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1";
RT Nucleic Acids Res. 28:862-867(2000).
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl
diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ010302; CAB38744.1; --

EMBL; AF195122; AAF24294.1; --

PIR; S49625; S49625.

PIR; T50750; T50750.

InterPro; IPR000092; Polyprenyl_synth.

InterPro; IPR008949; Terpenoid synth.

Pfam; PF00348; polyprenyl synth; 1.

PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.

PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.

Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;

Isoprene biosynthesis; Transferase.

CONFLICT 53 53 A -> R (IN REF. 1).

CONFLICT 57 57 D -> S (IN REF. 1).

CONFLICT 60 60 V -> A (IN REF. 1).

CONFLICT 123 123 R -> G (IN REF. 1).

CONFLICT 128 128 Q -> R (IN REF. 1).

CONFLICT 234 234 N -> S (IN REF. 1).

CONFLICT 247 247 R -> G (IN REF. 1).

CONFLICT 264 264 G -> A (IN REF. 1).

CONFLICT 277 277 E -> D (IN REF. 1).


```
SQ SEQUENCE 288 AA; 30256 MW; 5A7BDEBC41811A36 CRC64;
Query Match 20.1%; Score 307; DB 1; Length 288;
Best Local Similarity 35.5%; Pred. No. 1.3e-17;
Matches 88; Conservative 32; Mismatches 96; Indels 32; Gaps 8;
QY 43 ANRGDTLAP-----GRIIPMLLLTARDLGCAISHGGLDLAC-AVEKWH 87
DB 16 ARGQSEAPSKLATALDYAVTPGGARIPTLLSVA--TACGDDRPVLSDAVALELIH 73
QY 88 AASLILDDPCCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGVIAEAGLTPIAKT 147
DB 74 CASLHVHDLPCFDDAEIRRGRTVTHRAISEPLAALTGDSLWGFVLAARAADQFORAL 133
QY 148 RAVSELSTAIGM-QGLVGOQFQDLSEGDKPRSDAAILLTNQFKTSTLFCASTQMASIA 206
DB 134 QLVTALAVRTGMPMGICAGQ-----GWESQINLSAYHRAKTGALFIATQMGATAAG 187
QY 207 ASCEARENLHRSFLDLGQAFOLLDLTDMGTD--GKIDNDA--GKSTIVNLGSGAV 261
DB 188 YEAEFWEL---GARIGAFQVADDLRLDLCDAETLKGPAQODEIHARPNVREYGVEGA 244
QY 262 EERLRQHL 269
DB 245 AKRLKDL 252

RESULT 15
ISPA_MICLU STANDARD; PRT; 291 AA.
AC 066126;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)
DE (GPP synthase).
GN FPS.
OS Micrococcus luteus (Micrococcus lysodeikticus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Micrococcaceae; Micrococcus.
OX NCBI_TaxID=1270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-P 26;
RX MEDLINE=98175686; PubMed=9515931;
RA Shimizu N., Koyama T., Ogura K.;
RT "Molecular cloning, expression, and characterization of the genes
RT encoding the two essential protein components of Micrococcus luteus
RT B-P 26 hexaprenyl diphosphate synthase.";
RL J. Bacteriol. 180:1578-1581(1998).
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
CC = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the FPS/GGPP synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB003187; BAA25265.1; -
CC InterPro; IPR000092; Polyprenyl_synt.
CC InterPro; IPR008949; Terpenoid_synth.
CC Pfam; PF00348; polyprenyl_synt; 1.
CC PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
CC PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
CC Transferrase, isoprene biosynthesis.
SQ SEQUENCE 291 AA; 32362 MW; 797125AB71E5674A CRC64;

Query Match 20.0%; Score 304.5; DB 1; Length 291;
```

```
Best Local Similarity 36.2%; Pred. No. 2e-17;
Matches 92; Conservative 36; Mismatches 109; Indels 17; Gaps 7;
QY 24 DSRLLDLLPVGERDCVGAAMREGTLAPGKRIRPMLLLTARDLGCAISHGGLDLACAV 83
DB 18 ESLNRYHPAGSR---LHEAINYSLSAGGKRIRPLVLVTTLDLSLG-GNAHDG-LPFGIAL 72
QY 84 EMVHAASLILDDPCCMDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGVIAEAGLTPI 143
DB 73 EMHTYSLIHDDLPAAMDNDYRRGKLTNHRKFDDEATAILAGDALLTDAFOCILNTQ-LNA 131
QY 144 IAKTRAVSELSTAIGMQLVGOQFQDLSEGDKPRSDAAILLTNQFKTSTLFCASTQMASI 203
DB 132 EIKLSLIMLLSTAGSNGMVTYQQLDMQGEHKTUTLNELEERIHKHTGELTRAAIVSAGI 191
QY 204 AANASCEARENLHRSFLDLGQAFOLLDLT-----GMTDTGKDINQDAGKSTLVNL 255
DB 192 INWFDAQIEQLNIIGKNVGLMFMQIKDDILDVEGSPENIGKT-VGSDLNND--KSTTVSL 248
QY 255 LGSGAVEERLRQHL 269
DB 249 LGLRASKQLLNDKL 262

Search completed: February 29, 2004, 14:45:03
Job time : 6.10674 secs
```

GeoCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

in on: February 29, 2004, 14:51:24 ; Search time 20.1175 Seconds
(without alignments)

3180.293 Million cell updates/sec

File: US-09-941-947A-26

irfect score: 1526

quence: 1 LTVCAKXVHLTGISABQLL.....HSTQLFIQAFWFKLAAYS 303

oring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 211153259 residues

al number of hits satisfying chosen parameters: 809742

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09A_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1526	100.0	303	10	US-09-941-947A-26
2	1526	100.0	303	14	US-10-218-118-2
3	427	28.0	295	14	US-10-166-225A-185
4	425	27.9	295	9	US-09-547-267-1
5	425	27.9	295	10	US-09-920-923-2
6	412.5	27.0	294	15	US-10-369-493-17358
7	394.5	25.9	292	15	US-10-369-493-10617
8	388	25.4	269	13	US-10-108-915-18
9	381.5	25.0	294	15	US-10-369-493-18010
10	380.5	24.9	291	15	US-10-369-493-8506
11	376.5	24.7	287	9	US-09-925-637-64
12	376.5	24.7	287	14	US-10-084-205-64
13	375.5	24.6	293	9	US-09-815-242-12583
14	375.5	24.6	293	10	US-09-920-923-37
15	368.5	24.1	288	9	US-09-815-242-5239

Sequence 16943, A
Sequence 12014, A
Sequence 19922, A
Sequence 11449, A
Sequence 14816, A
Sequence 15003, A
Sequence 2608, Ap
Sequence 10187, A
Sequence 18752, A
Sequence 2, Appli
Sequence 20893, A
Sequence 23201, A
Sequence 4670, Ap
Sequence 7429, Ap
Sequence 45, Appl
Sequence 16549, A
Sequence 22, Appl
Sequence 14, Appl
Sequence 72, Appl
Sequence 20, Appl
Sequence 10069, A
Sequence 732, App
Sequence 15817, A
Sequence 16200, A
Sequence 15449, A
Sequence 19191, A
Sequence 9546, Ap
Sequence 14084, A
Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-941-947A-26
; Sequence 26, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picstagio, Steve
; APPLICANT: Roviore, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 26
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Pantoea stewartii
US-09-941-947A-26

Query Match 100.0%; Score 1526; DB 10; Length 303;
Best Local Similarity 100.0%; Pred. No. 2.3e-150;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LTVCAKXVHLTGISAEQLADISRLDQLLPVQGRDCVGAAMEGTLAPKRIKRPMLL 60
Db 1 LTVCAKXVHLTGISAEQLADISRLDQLLPVQGRDCVGAAMEGTLAPKRIKRPMLL 60
Qy 61 LITARDLGCAISHGGLLOLACAVEMVHAASLILDDMPCNDQAMRGRTTHTQYGEHVA 120

[illegible]

```

RESULT 2
JS-10-218-118-2
Sequence 2, Application US/10218118
Publication No. US20030148319A1
GENERAL INFORMATION:
APPLICANT: Brzostowicz, Patricia
APPLICANT: Rouviere, Pierre
APPLICANT: Picataggio, Stephen
APPLICANT: Cheng, Qiong
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds
FILE REFERENCE: CL1876 US NA
CURRENT APPLICATION NUMBER: US/10/218,118
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: 60/312,646
PRIOR FILING DATE: 2001-08-15
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 303
TYPE: PRT
ORGANISM: Pantoea stewartii
JS-10-218-118-2

Query Match 100.0%; Score 1526; DB 14; Length 303;
Best Local Similarity 100.0%; Pred. No. 2.3e-150;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 3

```

US-10-166-225A-185
? Sequence 185, Application US/10166225A
? Publication No. US2003014816A1
? GENERAL INFORMATION:
? APPLICANT: BERRY, Alan
? APPLICANT: HUMBELIN, Markus
? APPLICANT: LOPEZ-ULIBARRI, Rual
? APPLICANT: MAYER, Anne F.
? APPLICANT: YELISEEV, Alexei A.
? TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
? FILE REFERENCE: C38435/121966
? CURRENT APPLICATION NUMBER: US/10/166,225A
? CURRENT FILING DATE: 2002-06-05
? NUMBER OF SEQ ID NOS: 197
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 185
? LENGTH: 295
? TYPE: PRT
? ORGANISM: Paracoccus sp. RI534
US-10-166-225A-185

```

Query Match	28.0%; Score 427; DB 14; Length 295;
Best Local Similarity	39.6%; Prad. No. 6.8e-36;
Matches 111; Conservative	37; Mismatches 102; Indels 30; Gaps 19;
QY	23 IDSRLDQLLPVQGERDCV-----GAAMREGTLPAGKRIIRPMLILLFARDLGCASHGG--- 75
Db	13 VEIRLAQ---ISGFGVVSAPLGAASDAALSPGKFRVILMLVAE-----SSGGVCD 63
QY	76 -LLDLACAVEMVHHAASILLDDMCMDDAQNRGRPTIHQYGEHVAILAAVALLSKAFGV 134
Db	64 AMVDAACAVEMVHHAASILLFDDMCMDDATRGQPATHVAGGEGRAVLGALITAEWRI 123
QY	135 IAAEAGLPTIAKTRAVSELSTAIGCMQGLVQGFQDLSEGDKPRSDAILLTMQFKTSLP 194
Db	124 LCEARGATPDQARLVASMSRANGPVCLCAGQDLDL--HAPKDAAGIEREQDLKTGVLF 180
QY	195 CASTQVASIAANASCHARENLHFSLDLGCAFQLLDLDTGMDT--TGKDINOD-AGKS 250
Db	181 VAGLEMLSIIKGLDKAETEQLMAFGQLGRVFSYDLDLDVIGDKASTKGTGRTDTAPG 240
QY	251 TLVNLGSGAVEERLQRHLRLASEHLSAACQNGHSTOLF 290
Db	241 PKBGTLMVWGMD--VACHYASRAQGLELMR-----TRLF 274

RESULT 4
 US-09-547-267-1
 Sequence 1, Application US/09547267
 Patent No. US20020147371A1
 GENERAL INFORMATION:
 APPLICANT: Hohmann, Hans-Peter
 APPLICANT: Pasauntes, Luis
 APPLICANT: Tessier, Michel
 APPLICANT: van Loon, Adolphus
 TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Hoffmann-La Roche Inc.
 STREET: 340 Kingeland Street
 CITY: Nutley
 STATE: NJ
 COUNTRY: USA
 ZIP: 07110
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/547,267
 FILING DATE:

: APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Chen, Xianfeng
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 : FILE REFERENCE: 38-10152052/B
 : CURRENT APPLICATION NUMBER: US/10/369,493
 : CURRENT FILING DATE: 2003-02-28
 : PRIOR APPLICATION NUMBER: US 60/360,039
 : PRIOR FILING DATE: 2002-02-21
 : NUMBER OF SEQ ID NOS: 47374
 : SEQ ID NO 10617
 : LENGTH: 292
 : TYPE: PRT
 : ORGANISM: Sphingomonas aromaticivorans
 : IS-10-369-493-10617

Query Match	25.9%	Score 394.5	DB 15	Length 292
Best Local Similarity	36.0%	Pred. No. 1.6e-32		
Matches 95	Conservative 45	Mismatches 115	Indels 9	Gaps 5
22	DIISRLLDQLLPVOGE--RDCVGAAMRGSTLAPGKIRPMLLLLTARDLGAICASHGGGLDLA	80		
11	DIDESFDLLLPVPCDPEDRLVREMYAAIGGGKRLAPLLVCATGALHG--VYSREAARVVG	68		
81	CAVEMVIAASLLDDPCMDDDAQMWRGRPTIHTQYGEHVAIIAAVALLSKAPGVIAE--AE	139		
69	TAIEAIHVYSLIHDDLPQMDNDGLRGKPTVHVAPDDATAVLADGALLHDFEVLSDPAT	128		
140	GLTIETAKTRAVSELSTAIAGVQGLVQGFQDLSEGDKPRSDAILLUNQFTSTLFCASTQ	199		
129	SGDPFTRIELIRTLLTATASGNNCGGQMMDIVAETSSFDLPTVTRLOQLTKTGALIGASVE	188		
200	MASIAAANASCARENLERPSLDIGQAFQLDDLTQDQWTD---TGKDINDQA--GKSTLVN	254		
189	LGAILGKVPPEGRTHLRGYARDIGLAFQIADLLDLAGEDAAAGKALKRDKDAQGKETFLS	248		
255	LLGSGAVEERLRQHLRLASEHLSA	278		
249	LLGADRAREQARLLVNOAVEHLAS	272		

```

RESULT 6
US-10-108-915-18
; Sequence 18, Application US/10108915
; Publication No. US20020177204A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Shen, Jennie
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: Geranylgeranyl Pyrophosphate Synthases
; FILE REFERENCE: BHI286 US NA
; CURRENT APPLICATION NUMBER: US/10/108,915
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/452,238
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/110,592
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Glycine max
US-10-108-915-18

```

Query Match 25.4%; Score 388; DB 13; Length 369;
Best Local Similarity 37.3%; Pred. No. 1.1e-31;
Matches 98; Conservative 49; Mismatches 104; Indels 12; Gaps 6;

```

DB      89 VNQALDAIAIALRDPHK-IHQMRYSILAGGRKRPVLCIACELVGG--TEATAIPAACA 145
QY      83 VEMVHASLIILDDMPQMDDAQMRGRPTIHTQYGEHVAIIAAVALLSKAF-GVTAIEAEG 141
DB      146 VEMIHTMSLIHDDLPQMDNDLRRGKPTNHKVYGEDVAVLAGDALLAFEFHVAASTEGV 205
QY      142 TPIAKTRAVSELSAINGOVLQVQOPKDL-SEGDKPRSAIDAIIITNPKSTLFCASTQM 200
DB      206 SPSRVVRAIGELAKSIGTEGLVAGQVDDISEGVANVGLTLEFIHVHKTAALLERAAVVL 265
QY      201 ASIAANASCEARENLRHFSLDLGAQFQLLDLTGQMTDTGKDINQDRGKS-----TLVN 254
DB      266 GAIVGGSGDSEVEKLRKPARCIGLLQFQVDDILD-VTKSSELGKTAGKDLVADKVTPK 324
QY      255 ILGSGAVERLRQHLRLASEHLS 277
DB      325 LLGIDKSKFEAQELLKDAKQSLS 347

RESULT 9
US-10-369-493-18010
; Sequence 18010, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18010
; LENGTH: 294
; TYPE: PRT
; ORGANISM: SPHINGOMONAS
US-10-369-493-18010

```

Query Match	25.0%	Score 381.5;	DB 15;	Length 294;
Best Local Similarity	36.9%;	Prod. No. 3,7e-31;		
Matches 101;	Conservative 42;	Mismatches 110;	Indels 21;	Gaps 9;
Qy	18	QLLADIDSELLOLLPVQGE-RDCVGAMREGTLPAGKRIRPMLLLLTAR----	DIGCAIS	72
Db	10	QVAAEIDQDFRLLEVPDDPRADILYRAMHAAIGGSKRLPILVFETAQLFSDVKSCAAR	69	
Qy	73	HGGLLDLDAVENVHAAALLIDMPCWDDAQRGEPETHTOYGEHVAILAAVALLSKAF	132	
Db	70	-----VATALECHIVYSLIHDPLPAMDDDMRGKPTVHKAFDSEATILAGCDLHDLAF	123	
Qy	133	GVIA--EABGLTPIAKTRAVSELSTAIGMQLVGQGFDSGDKPRSDAILLT--NOFK	189	
Db	124	ELLAHEATHADPFVVRVELISELARSSCPAGMAGGQWML-EAEK-ASFDFLATVTFLOALK	181	
Qy	190	TSITLFCASTOMASTAANASCEARENHRESLDLGGAFQLDLDTGMDT---TGKNDINQD	246	
Db	192	TGALISCAVECGAILGRVPHGRTGLRGARDLGLAFQIVDDILDAEGDEAVVGKLGKD	241	
Qy	247	--AGKSTLVNLLGSAVEERLRQHLRLASEHLSA	278	
Db	242	GAAGKETLSILIGDRAREQARMLVQQAHLHA	275	

RESULT 10
US-10-369-493-8506
; Sequence 8506, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12583
LENGTH: 293
TYPE: PRT
ORGANISM: Staphylococcus aureus

Query Match 24.6%; Score 375.5; DB 9; Length 293;
Best Local Similarity 35.6%; Pred. No. 1.6e-30;
Matches 98; Conservative 43; Mismatches 119; Indels 15; Gaps 5;
Y 11 LTGASQLLADISRLDQLLPVQGERDCVGAANREGTL-----APKRIKRPMLLLTARD 66
b 1 MTNLPNNKLIDVNNLS-----VAINKSVMDTQLEBSMLYSLNAGGKIRPVLILLTLD 56
Y 67 LGCAISHGGLDLACAVEMVHAASLIIDDPCKDDAOMRRGRPTIHTQYGEHVAILAAVA 126
b 57 LNTYEYLG--KMSAIALEMHTYSLIHDDLPAMDNDYRGKLTNKHVYGEVAILAGDA 114
Y 127 LLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDAILLTN 186
b 115 LTKAFELISSDRLTDEVKIKVLQRLSIASGHVGMVGGQMLDQSEGQFIDLETLEMT 174
Y 187 QFKTSTLFCASTOMASIAANASCEARENHRSFLDLGQAFQLDDLTGQMTD---TGKDI 243
b 175 KTKTGALLTFAMGAADIANVDDATKHELSYSYHLGMMFQIKDDLLDCYGEAKLGKV 234
Y 244 NQDA--GKSTLVNLGSGAVEERLQHLRLASEHL 276
b 235 GSDLENNKSTVSLGKGABDKLTYHRDAVDEL 269

RESULT 14
US-09-920-923-37
Sequence 37, Application US/09920923
Publication No. US2003002273A1
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Improved Fermentative Carotenoid Production
FILE REFERENCE: Improved Fermentative Carotenoid
CURRENT APPLICATION NUMBER: US/09/920,923
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 56

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 37
LENGTH: 293
TYPE: PRT
ORGANISM: E-396
OTHER INFORMATION: Description of Unknown Organism: Unknown
US-09-920-923-37
Query Match 24.6%; Score 375.5; DB 10; Length 293;
Best Local Similarity 38.3%; Pred. No. 1.6e-30;
Matches 90; Conservative 38; Mismatches 86; Indels 21; Gaps 6;
Y 23 IDRLDQLLPVQ---ERDCVGAANREGTLAPKRIKRPMLLLTARDLGCALSHGG--- 75
b 13 LQTLLEI--AQGFVAVSOPFGPMGSHGALSCKRFRGMLMLAAE-----ASGGVCDT 64
Y 76 LLDLACAVEMVHAASLIIDDPCKDDAOMRRGRPTIHTQYGEHVAILAAVALLSKAPGVI 135
b 65 IVDAAVEMVHAASLIIDDPCKDDAOMRRGRPTIHTQYGEHVAILAAVALLSKAPGVI 124
Y 136 ABAEGLTPIAKTRAVSELSTAIGMGLVQGFQKDLSEGDKPRSDAILLTNPKTSTLFC 195
b 125 AGARGASGTVAQLVRLSRSLGPOGLCAQDLDLHAA---KNGAGVEQEDLKTGVLFI 181
Y 196 ASTOMASIAANASCEARENHRSFLDLGQAFQLDDLTGQMTD---TGKDIHQDA 247
b 182 AGLEMLAVIKFPAERGTQMDIFGRQLGRVFSYDLDLVGQDAALGKDTGRDA 236

RESULT 15
US-09-815-242-5239
Sequence 5239, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5239
LENGTH: 288
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5239
Query Match 24.1%; Score 368.5; DB 9; Length 288;
Best Local Similarity 36.2%; Pred. No. 8.1e-30;
Matches 97; Conservative 41; Mismatches 115; Indels 15; Gaps 5;

18 QLLADIDSRLDQLLEVOGERDCVGAAMEGTL-----APGKRIKIRPMMLLLLTARDLGCAISH 73
3 KLIDEVNNELS-----VAINKSVMDTQLEESMLYSINAGGKIRPVLTLTLDLSLNTYEL 58
74 GCLLDLACAVENWHAASJLDDMPCMDDAQMERGPTHTOYGEHVALLAAVALLSKAFG 133
59 G--MKSATALEMIHTYSLIHDLPAMDNDYRRGKLTNHKVIKVTAILAGDALLTKAFE 116
134 VIAEAGLTPIAKTRAVSELSTAIGMQLVOGQFQKLSGDKPRSDAAILITNOFKTSTL 193
117 LSSDDRLTDEVKIKVQLSLIASGHVGVGGQMLDMQSEGQPIDLEMLEMIHKTGTAL 176
194 FCASQMASIAANASCBARENLRFSLDLGOAFQLLDDLTGMDT---TGKDINQDA--G 248
177 LTFVMSAADIANVDDATKEHLESYSYHLGMVFOIKDLDLCYGDGAKLGKKVGSLENN 236
249 KSTLVNLLGSGAVEERLRQHLRLASEHL 276
237 KSTYVSLLGKGAEDKLTYYHRDAVDEL 264

Search completed: February 29, 2004, 15:27:54
Job time : 22.1175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 9.20761 Seconds
(without alignments)
1698.885 Million cell updates/sec

itle: US-09-941-947A-26

fect score: 1526
quence: 1 LTVCKKHVHTGISAEQLL.....HSTQLFIQWFDKLAAYS 303

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1344.5	88.1	302	1 US-07-783-705A-1	Sequence 1, Appli
2	763.5	50.0	298	1 US-08-095-725-4	Sequence 4, Appli
3	763.5	50.0	298	1 US-08-096-043-4	Sequence 4, Appli
4	763.5	50.0	298	1 US-08-093-577-4	Sequence 4, Appli
5	763.5	50.0	298	1 US-08-096-623A-4	Sequence 4, Appli
6	763.5	50.0	307	1 US-08-095-726-2	Sequence 2, Appli
7	763.5	50.0	307	1 US-08-096-043-2	Sequence 2, Appli
8	763.5	50.0	307	1 US-08-093-577-2	Sequence 2, Appli
9	763.5	50.0	307	1 US-08-096-623A-2	Sequence 2, Appli
10	425	27.9	295	3 US-08-660-645A-1	Sequence 1, Appli
11	425	27.9	295	3 US-09-298-718-1	Sequence 1, Appli
12	425	27.9	295	3 US-09-546-969-1	Sequence 1, Appli
13	425	27.9	295	3 US-08-980-832-2	Sequence 2, Appli
14	425	27.9	295	4 US-09-547-267-1	Sequence 2, Appli
15	425	27.9	295	4 US-09-920-923B-2	Sequence 2, Appli
16	375.5	24.6	293	3 US-08-980-832-37	Sequence 37, Appli
17	375.5	24.6	293	3 US-09-920-923B-37	Sequence 37, Appli
18	367.5	24.1	293	3 US-09-276-873-2	Sequence 2, Appli
19	360.5	23.6	297	1 US-08-534-910B-8	Sequence 8, Appli
20	360.5	23.6	297	3 US-08-886-466-2	Sequence 2, Appli
21	360.5	23.6	297	3 US-09-475-304-2	Sequence 2, Appli
22	360.5	23.6	297	4 US-09-101-126-3	Sequence 3, Appli
23	360.5	23.6	297	4 US-09-367-528A-5	Sequence 5, Appli
24	359.5	23.6	297	1 US-08-534-910B-7	Sequence 7, Appli
25	359.5	23.5	297	1 US-08-534-910B-9	Sequence 9, Appli
26	358.5	23.5	297	1 US-08-534-910B-10	Sequence 10, Appli
27	357.5	23.4	285	3 US-09-187-050-12	Sequence 12, Appli

28	357.5	23.4	393	3 US-09-187-050-2	Sequence 2, Appli
29	357.5	23.4	393	3 US-09-187-050-14	Sequence 14, Appli
30	357.5	23.4	393	3 US-09-187-050-16	Sequence 16, Appli
31	357.5	23.4	393	3 US-09-187-050-18	Sequence 18, Appli
32	357.5	23.4	393	3 US-09-187-050-20	Sequence 20, Appli
33	357.5	23.4	393	3 US-09-187-050-22	Sequence 22, Appli
34	357.5	23.4	393	3 US-09-187-050-24	Sequence 24, Appli
35	357.5	23.4	393	3 US-09-187-050-26	Sequence 26, Appli
36	357.5	23.4	393	3 US-09-187-050-27	Sequence 27, Appli
37	357.5	23.4	393	3 US-09-187-050-28	Sequence 28, Appli
38	357.5	23.4	393	3 US-09-187-050-29	Sequence 29, Appli
39	357.5	23.4	393	3 US-09-187-050-30	Sequence 30, Appli
40	357.5	23.4	393	3 US-09-187-050-31	Sequence 31, Appli
41	357.5	23.4	393	3 US-09-187-050-32	Sequence 32, Appli
42	357.5	23.4	393	3 US-09-187-050-33	Sequence 33, Appli
43	357.5	23.4	393	3 US-09-187-050-34	Sequence 34, Appli
44	356.5	23.4	297	4 US-09-367-528A-1	Sequence 1, Appli
45	354.5	23.2	297	1 US-08-534-910B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-07-783-705A-1
; Sequence 1, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOID
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Ladas & Parry
; STREET: 26 West 61 Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/783,705A
; FILING DATE: 19911023
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-103078
; FILING DATE: 21-APR-1989
; APPLICATION NUMBER: JP 2-53225
; FILING DATE: 05-MAR-1990
; APPLICATION NUMBER: US 07/519,011
; FILING DATE: 13-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwadron, Janet I.
; REGISTRATION NUMBER: 33,778
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-708-1935
; TELEFAX: 212-246-5959
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-783-705A-1
Query Match 88.1%; Score 1344.5; DB 1; Length 302;

Query Match	50.08;	Score 763.5;	DB 1;	Length 298;
Best Local Similarity	56.08;	Pred. No. 1.5e-79;		
Matches 159;	Conservative 40;	Mismatches 84;	Indels 1;	Gaps 1;
17	EQLLADIDSRLDQLLPYQGEDCVGAAMRGTTAPGKRIKRPMLLLLTARDLGCASHHGL	76		
7	EVMRQSIDDHLAGLIPPTDSDIVSLAMRGVWAPGKRIKRPMLLLLAARDLRYGSMPTL	66		
77	LDLCAVEMVHAASLIIDDMPCMDAOWMRGRPTHTFOYGEHVAILAAVALLSKAFGVIA	136		
67	LDLCAVELTHTASEMLDDMPMDNAELRGQP:THKKFGESVAILASVGLLSKAPGLIA	126		
137	BAEGLTPIAKTRAVSEI:STAIGMGLVQGFQKLSGEGDKPSPADAILLTNQFTKSTL:FCA	196		
127	ATGDLPGERRAQA:VNEI:STAVLGQLVIGQFRLDNLDAALDRTPDAILSTNHLKTGIL:FSA	186		
197	STOWASIAANASCEARNLHRFSLDLGOAFQLLDDLTGKMTDGTGKINODAGKSTLVNLL	256		
187	MLQIVAIASASPSTRTHLHAFD:FQA:FQLLDDLRDDHPETGKDKAKGKSTLVNRL	246		
257	GSGAVEERLRQHLRLASEHLSAAQNGHSTTQLPFIQAWFNKLA	300		
247	GADAAKOKLRHHIOSAKUHLTFA:CPQGAIRQ:FMHLWFGHLLA	289		

RESULT 3
 S-08-096-043-4 Application US/08096043
 Sequence 4, Patent No. 5530189
 GENERAL INFORMATION:
 APPLICANT: Ausich, Rodney L
 APPLICANT: Brinkhaus, Friedhelm L
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H
 APPLICANT: Yarger, James G
 APPLICANT: Yen, Hwei-Che B
 TITLE OF INVENTION: Lycopen Biosynthesis in
 TITLE OF INVENTION: Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 70
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Amoco Corp., Patents and Licensing Dept
 STREET: 200 E Randolph St
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60680-0703
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096.043
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/785,568
 FILING DATE: 30-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Galloway, No. 5530189val B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3128567180
 TELEFAX: 3128564972
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 298 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 S-08-096-043-4

Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

17 EQLADIDSRLDQLPVQGRDCVGAAMREGTLAPGKRIRPMLLLLTARDLGCASHGGL 76
7 EVMRQSIDDHLAGLLPETSQDIVSLAMREGVWAPGKRIRPLMLLAARDLRYQGSMPYL 66
77 LDLCACAVEMVHAASLILDDMPQMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAPGVIA 136
67 LDLCACAVELTHTASLMLDDMPQMDNAELRRGQPTTHKTFGESVAILASVGLLSKAPGLIA 126
137 EASGLTPIAKTRAVSELSTAIGMQGLVQGFQKLSGDKPRSDAILLTNQKFTSLFCA 196
127 ATGDLPGERRAQAVNELSTAVGLQGLVGFQFRDLNDAALDRTPDAILSTNHLKTGLFSA 186
197 STQMASTAAASCEARENLRHFSLDLQGFQFLLDLDLTGDMTGTGKDIINODACKSTLVNLL 256
187 MLQIVAIASASSPSTRETLHAFALDQGFQFLLDLDLDDHPETGKDRNKDACKSTLVNRL 246
257 GSGAVERLEBOHLRLASEHLSAACQNGHSTTQLFQIAWFDKILA 300
247 GADARQKUREHIDSADKHLTFACPOGGAIRO-FMELMFGHILA 289

```

RESULT 4

-08-093-577-4
Sequence 4, Application US/08093577
Patent No. 5545816

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukhartji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Phycoene Biosynthesis in
Genetically Engineered Hosts
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093,577
FILING DATE: 19-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,569
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5545816val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

-08-093-577-4
Query Match 50.0%; Score 763.5; DB 1; Length 298;
Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

17 EQLADIDSRLDQLPVQGRDCVGAAMREGTLAPGKRIRPMLLLLTARDLGCASHGGL 76
7 EVMRQSIDDHLAGLLPETSQDIVSLAMREGVWAPGKRIRPLMLLAARDLRYQGSMPYL 66
77 LDLCACAVEMVHAASLILDDMPQMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAPGVIA 136
67 LDLCACAVELTHTASLMLDDMPQMDNAELRRGQPTTHKTFGESVAILASVGLLSKAPGLIA 126
137 EASGLTPIAKTRAVSELSTAIGMQGLVQGFQKLSGDKPRSDAILLTNQKFTSLFCA 196
127 ATGDLPGERRAQAVNELSTAVGLQGLVGFQFRDLNDAALDRTPDAILSTNHLKTGLFSA 186
197 STQMASTAAASCEARENLRHFSLDLQGFQFLLDLDLTGDMTGTGKDIINODACKSTLVNLL 256
187 MLQIVAIASASSPSTRETLHAFALDQGFQFLLDLDLDDHPETGKDRNKDACKSTLVNRL 246
257 GSGAVERLEBOHLRLASEHLSAACQNGHSTTQLFQIAWFDKILA 300
247 GADARQKUREHIDSADKHLTFACPOGGAIRO-FMELMFGHILA 289

```

Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

17 EQLADIDSRLDQLPVQGRDCVGAAMREGTLAPGKRIRPMLLLLTARDLGCASHGGL 76
7 EVMRQSIDDHLAGLLPETSQDIVSLAMREGVWAPGKRIRPLMLLAARDLRYQGSMPYL 66
77 LDLCACAVEMVHAASLILDDMPQMDDAQMRGRPTIHTQYGEHVAILAAVALLSKAPGVIA 136
67 LDLCACAVELTHTASLMLDDMPQMDNAELRRGQPTTHKTFGESVAILASVGLLSKAPGLIA 126
137 EASGLTPIAKTRAVSELSTAIGMQGLVQGFQKLSGDKPRSDAILLTNQKFTSLFCA 196
127 ATGDLPGERRAQAVNELSTAVGLQGLVGFQFRDLNDAALDRTPDAILSTNHLKTGLFSA 186
197 STQMASTAAASCEARENLRHFSLDLQGFQFLLDLDLTGDMTGTGKDIINODACKSTLVNLL 256
187 MLQIVAIASASSPSTRETLHAFALDQGFQFLLDLDLDDHPETGKDRNKDACKSTLVNRL 246
257 GSGAVERLEBOHLRLASEHLSAACQNGHSTTQLFQIAWFDKILA 300
247 GADARQKUREHIDSADKHLTFACPOGGAIRO-FMELMFGHILA 289

```

RESULT 5

US-08-096-623A-4
Sequence 4, Application US/08096623A
Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukhartji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
Glycosylated Zeaxanthin in Genetically Engineered Hosts
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Welsh & Katz, Ltd.
STREET: 120 S. Riverside Plaza, 22nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,623A
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,061
FILING DATE: 09-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,921
FILING DATE: 28-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/562,674
FILING DATE: 03-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/525,551
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/487,613
FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500

TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 amino acids
TYPE: amino acid
TOPOLOGY: linear
STRANDEDNESS:
MOLECULE TYPE: protein
US-08-095-726-2

Query Match 50.0%; Score 763.5; DB 1; Length 298;
Best Local Similarity 56.0%; Pred. No. 1.5e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
Y 17 EQLADIDSRDLQVQGERDCVGAAMEGRTIAPGKIRPMLLLLTARDLGCALSHGGL 76
b 7 EVNRQSIDHLAGLLPETUSQDIVSLAMEGVMAFGKIRPLMLLLAARDLRYQGSMTPL 66
Y 77 LDACAVEMVHAASLILDDMPQVDDQAMRGRTTHQYGEHVAILAAVALLSKAFGVIA 136
b 67 LDACAVELTHTASLMLDDMPQVDDQAMRGRTTHQYGEHVAILAAVALLSKAFGLIA 126
Y 137 EAGLTPIAKTRAVSELSTAIGMVGQVQFQDLSEGDKPRSDAILLTNPKTSTLPCA 196
b 127 ATGDLPGERRAQAVNELSTAVGLQGLVGFQDLNDALDRTPDAILSTNHLKTLGILFSA 186
Y 197 STQMASIAANASCEARENHRFSLDLQAFQQLDDLTGQMTDTGKIDNODACKSTLVNLL 256
b 187 MLQIVAIASASSPSTETLHAFALDFQAFQQLDDLRDDHPETGKRNKDGKSTLVNRL 246
Y 257 GSCAVERLRQHLRLASEHLSAACQNGHSTTQLFQAMFDKCLA 300
b 247 GADAAARQKLEHIDSADKHLTFACPOGGAIRQ-FMHLWFGHLLA 289

ESULT 6

S-08-095-726-2
Sequence 2, Application US/08095726
Patent No. 5530188
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Beta-Carotene Biosynthesis in
Genetically Engineered Hosts
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530188val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-726-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
Best Local Similarity 56.0%; Pred. No. 1.6e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
QY 17 EQLADIDSRDLQVQGERDCVGAAMEGRTIAPGKIRPMLLLLTARDLGCALSHGGL 76
DB 16 EVNRQSIDHLAGLLPETUSQDIVSLAMEGVMAFGKIRPLMLLLAARDLRYQGSMTPL 75
QY 77 LDACAVEMVHAASLILDDMPQVDDQAMRGRTTHQYGEHVAILAAVALLSKAFGVIA 136
DB 76 LDACAVELTHTASLMLDDMPQVDDQAMRGRTTHQYGEHVAILAAVALLSKAFGLIA 135
QY 137 EAGLTPIAKTRAVSELSTAIGMVGQVQFQDLSEGDKPRSDAILLTNPKTSTLPCA 196
DB 136 ATGDLPGERRAQAVNELSTAVGLQGLVGFQDLNDALDRTPDAILSTNHLKTLGILFSA 195
QY 197 STQMASIAANASCEARENHRFSLDLQAFQQLDDLTGQMTDTGKIDNODACKSTLVNLL 256
DB 196 MLQIVAIASASSPSTETLHAFALDFQAFQQLDDLRDDHPETGKRNKDGKSTLVNRL 255
QY 257 GSCAVERLRQHLRLASEHLSAACQNGHSTTQLFQAMFDKCLA 300
DB 256 GADAAARQKLEHIDSADKHLTFACPOGGAIRQ-FMHLWFGHLLA 298

RESULT 7

US-08-096-043-2
Sequence 2, Application US/08096043
Patent No. 5530189
GENERAL INFORMATION:
APPLICANT: Ausich, Rodney L
APPLICANT: Brinkhaus, Friedhelm L
APPLICANT: Mukharji, Indrani
APPLICANT: Proffitt, John H
APPLICANT: Yarger, James G
APPLICANT: Yen, Hwei-Che B
TITLE OF INVENTION: Lycopene Biosynthesis in
Genetically Engineered Hosts
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corp., Patents and Licensing Dept
STREET: 200 E Randolph St
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,043
FILING DATE: 22-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,568
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5530189val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 307 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 3-08-096-043-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
 Best Local Similarity 56.0%; Pred. No. 3.6e-79;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

17 EQLADIISRLDQLPVQGERDVGAMREGTLPAGKRIIRPMLLLTARDLGCAISHGGL 76
16 EVMRSIDHLAGLIPETDSQDIVSLAWREGVMAFGKRIIRPMLLLAARDLRYQGSMTL 75
77 LDLACAVEMVHAASILLDDMPCMDAQMRRGPTTHQYGEHVAILAAVALLSKAFGVIA 136
76 LDLACAVELTHFASLMLDMPMDNAELRRGPTTHKXFGESVAILASVGLLSKAFGLIA 135
137 EAEGLTPIAKTAVSELSTAIGMQLVQGFQDLSEGDKPRSAADAILLTNQFKSTLPCA 196
136 ATGDLPGERRAQNVELSTAVGLQGLVGFQDLNDALDRTPDAILSTNHLKTGILFSA 195
197 STOMASIAANASCEARENHLRPSLDLGOAFQLLDLDLTDGMDTDTGKDINQDAGKSTLVNLL 256
196 MLQIVAIASASPSSTRETHAFALDFGOAFQLLDLDLDDHDPETGKRNKDGAKSTLVNRL 255
257 GSGAVEERLROHLRLASHELSAACQNGHSTTQLFTQAWFDKCLA 300
256 GADAAQKRLREHIDSADKELTPACPGGAIRQ-FMHLWFGHHLA 298
  
```

RESULT 8

3-08-093-577-2
 Sequence 2, Application US/080935577
 Patent No. 5545816

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.
 APPLICANT: Brinkhaus, Friedhelm L.
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H.
 APPLICANT: Yarger, James G.
 APPLICANT: Yen, Hui-Che B.
 TITLE OF INVENTION: Phytoene Biosynthesis in
 Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Amoco Corp., Patents and Licensing Dept
 STREET: 200 E Randolph St
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60680-0703

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/093,577
 FILING DATE: 19-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/785,569
 FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:
 NAME: Galloway, No. 5545816val B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 3128567180
 TELEFAX: 3128564972
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 307 amino acids
 TYPE: amino acid

TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-093-577-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
 Best Local Similarity 56.0%; Pred. No. 1.6e-79;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

```

17 EQLADIISRLDQLPVQGERDVGAMREGTLPAGKRIIRPMLLLTARDLGCAISHGGL 76
16 EVMRSIDHLAGLIPETDSQDIVSLAWREGVMAFGKRIIRPMLLLAARDLRYQGSMTL 75
77 LDLACAVEMVHAASILLDDMPCMDAQMRRGPTTHQYGEHVAILAAVALLSKAFGVIA 136
76 LDLACAVELTHFASLMLDMPMDNAELRRGPTTHKXFGESVAILASVGLLSKAFGLIA 135
137 EAEGLTPIAKTAVSELSTAIGMQLVQGFQDLSEGDKPRSAADAILLTNQFKSTLPCA 196
136 ATGDLPGERRAQNVELSTAVGLQGLVGFQDLNDALDRTPDAILSTNHLKTGILFSA 195
197 STOMASIAANASCEARENHLRPSLDLGOAFQLLDLDLTDGMDTDTGKDINQDAGKSTLVNLL 256
196 MLQIVAIASASPSSTRETHAFALDFGOAFQLLDLDLDDHDPETGKRNKDGAKSTLVNRL 255
257 GSGAVEERLROHLRLASHELSAACQNGHSTTQLFTQAWFDKCLA 300
256 GADAAQKRLREHIDSADKELTPACPGGAIRQ-FMHLWFGHHLA 298
  
```

RESULT 9

US-08-096-623A-2
 Sequence 2, Application US/08096623A
 Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.
 APPLICANT: Brinkhaus, Friedhelm L.
 APPLICANT: Mukharji, Indrani
 APPLICANT: Proffitt, John H.
 APPLICANT: Yarger, James G.
 APPLICANT: Yen, Hui-Che B.
 TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
 Glycosylated Zeaxanthin in Genetically Engineered Hosts
 NUMBER OF SEQUENCES: 104
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Welsh & Katz, Ltd.
 STREET: 120 S. Riverside Plaza, 22nd Floor
 CITY: Chicago
 STATE: IL
 COUNTRY: USA
 ZIP: 60606

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096,623A
 FILING DATE: 22-JUL-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,061
 FILING DATE: 09-DEC-1991

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/662,921
 FILING DATE: 28-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/562,674
 FILING DATE: 03-AUG-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/525,551
 FILING DATE: 18-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/487,613

FILING DATE: 02-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: AMO-006.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 655-1500
TELEFAX: (312) 655-1501
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-623A-2

Query Match 50.0%; Score 763.5; DB 1; Length 307;
Best Local Similarity 56.0%; Pred. No. 1.6e-79;
Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

QY 17 EQLADIISRLDQLLPVQGERDVCVAMREGLTAPGKRIKIRPMLLLTARDLGAISHGGL 76
DB 16 EVMQSIDHLAGLLETDSQDIIVSLAMREGVMAKRIKIRPMLLLAARDLRYQGSNPTL 75
QY 77 LELACAVEMVHAASLILDDPCMDDAQMRRCPTIHTQYGEHVAILAAVALLSKAFGVIA 136
DB 76 LELACAVELTHASLMLDDMFCMDNAELRRGQPTTHKKFGESVAILASVGLLSKAFGLIA 135
QY 137 EAGLTPTIATKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRASDAIILLTNQKTTSTLFLCA 196
DB 136 ATGDLPGERRAQAVNELSTAVGLQSLVQGFRLDNDAAALDTPDAIILSTHLLKGTGLFSA 195
QY 197 STOMASIAANASCEARENHLRFSLDLQAFQFOLLDDLTGMDTDTGKDINODAGKSTLVNML 256
DB 196 YLQIVAILASASPSTETILHAFALDFGQAFQFOLLDDLRDDHETGKRNKAGKSTLVNRL 255
QY 257 GSGAVEERLQHLRLASEHLSAACQNGHSTTOLFQIAWFDKLLA 300
DB 256 GADAAQKLRHIDSGAKHLTFACFGGAIRQ-FMLLWFGHLLA 298

RESULT 10
US-08-660-645A-1
Sequence 1, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-660-645A-1

Query Match 27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;
Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
QY 23 IDSRDLQQLPVQGERDCV---GAAMEEGTLAPGKRIKIRPMLLLTARDLGAISHG--- 75
DB 13 VEIRLAQ---ISGQGVVSAPLGAASDAALSPOKRFRAVLMMLVAE-----SSGGVCD 63
QY 76 -LLDLACAVEMVHAASLILDDPCMDDAQMRRCPTIHTQYGEHVAILAAVALLSKAPGV 134
DB 64 AMVDAACAVEMVHAASLIFDDPCMDDAQMRRCPTIHTQYGEHVAILAAVALLSKAPGV 123
QY 135 IARABGLTPIATKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRASDAIILLTNQKTTSTLP 194
DB 124 LGRAGATPDQARLIVASMSRAMGPVGLCAGDLDL---HAPKDAAGIEREQDLKTGVLF 180
QY 195 CASTOMASIAANASCEARENHLRFSLDLQAFQFOLLDDLTGMDTDTGKDINOD-AGKS 250
DB 181 VAGLEMLSIIRKGLDKAETEQLMFGRQGRVFGVYDLDLVIGDKASTGKDTARDTAPG 240
QY 251 TLVNLGSGAVEERLQHLRLASEHLSAACQNGHSTTOLF 290
DB 241 PKGGLMAYQMGD-VAQHYRASRAQLDELME-----TLFLP 274

RESULT 11
US-09-298-718-1
Sequence 1, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

-09-298-718-1

Query Match 27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSLDQLLPVQGERDCV---GAAMEGTLAPGKRIRPMLLLLTARDLGCALSHGG---75

13 VEIRLAQ---ISGGFGVVSAPFGAAMSDAALSPGKRFRAVLMVAE-----SSGGVCD 63

76 -LLDLACAVENVHAASLILDDMPQDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVENVHAASLIFDDMPQDDARTRRGQPATHVAHGEGRAVLGIALITEAMRI 123

135 IAEAGLTPIAKTRAVSELSTAICMQGLVQGFKDLSEGDKPSADAILLTNQFKTSTLF 194

124 LGARGATPDQARLVASMSRAMGPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLF 180

195 CASTOMASIAANASCEARENHRSFSLDLGQAFQLDLDTGMDT---TGKDINOD-AGKS 250

181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFSYDLDLVDIGDKASTGKTARDTAAPG 240

251 TLVNLGSGAVEERLRLHRLASEHLSAACQNGHSTTOLF 290

241 PKGGLMVGQMGD-VACHYRASRAQLDELNR-----TRLF 274

SUITE 12

-09-546-969-1

Sequence 1, Application US/09546969

Patent No. 6207409

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSES: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nucleu

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/546,969

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 295 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-546-969-1

Query Match 27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSLDQLLPVQGERDCV---GAAMEGTLAPGKRIRPMLLLLTARDLGCALSHGG---75

13 VEIRLAQ---ISGGFGVVSAPFGAAMSDAALSPGKRFRAVLMVAE-----SSGGVCD 63

76 -LLDLACAVENVHAASLILDDMPQDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVENVHAASLIFDDMPQDDARTRRGQPATHVAHGEGRAVLGIALITEAMRI 123

135 IAEAGLTPIAKTRAVSELSTAICMQGLVQGFKDLSEGDKPSADAILLTNQFKTSTLF 194

124 LGARGATPDQARLVASMSRAMGPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLF 180

195 CASTOMASIAANASCEARENHRSFSLDLGQAFQLDLDTGMDT---TGKDINOD-AGKS 250

181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFSYDLDLVDIGDKASTGKTARDTAAPG 240

251 TLVNLGSGAVEERLRLHRLASEHLSAACQNGHSTTOLF 290

241 PKGGLMVGQMGD-VACHYRASRAQLDELNR-----TRLF 274

RESULT 13

US-08-980-832-2

Sequence 2, Application US/08980832B

Patent No. 6231204

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/08/980,832B

CURRENT FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 2

LENGTH: 295

TYPE: PRT

ORGANISM: Flavobacterium sp. R1534

US-08-980-832-2

Query Match

27.9%; Score 425; DB 3; Length 295;

Best Local Similarity 39.6%; Pred. No. 1.8e-40;

Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;

23 IDSLDQLLPVQGERDCV---GAAMEGTLAPGKRIRPMLLLLTARDLGCALSHGG---75

13 VEIRLAQ---ISGGFGVVSAPFGAAMSDAALSPGKRFRAVLMVAE-----SSGGVCD 63

76 -LLDLACAVENVHAASLILDDMPQDDAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134

64 AMVDAACAVENVHAASLIFDDMPQDDARTRRGQPATHVAHGEGRAVLGIALITEAMRI 123

135 IAEAGLTPIAKTRAVSELSTAICMQGLVQGFKDLSEGDKPSADAILLTNQFKTSTLF 194

124 LGARGATPDQARLVASMSRAMGPVGLCAGQDLDL---HAPKDAAGIEREQDLKTGVLF 180

195 CASTOMASIAANASCEARENHRSFSLDLGQAFQLDLDTGMDT---TGKDINOD-AGKS 250

181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFQSYDDLLDVIGDKASTGKTARDTAAPG 240
251 TLVNLGSGAVERLRQHLRLASEHLSAACQNGHSTTOLF 290
241 PKGLMAVGQMGD-VAQHYRASRAQLDELNR-----TRLF 274

RESULT 14
US-09-547-267-1
Sequence 1, Application US/09547267
Patent No. 6613543
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-peter
APPLICANT: Passamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Ioon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/547,267
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-547-267-1
Query Match 27.9%; Score 425; DB 4; Length 295;
Best Local Similarity 39.6%; Pred. No. 1.8e-40;
Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
23 IDSRDLQLLPVQGERDCV-----GAAMREGTLAPKRIKIRPMLLLLTARDLGCALSHGG--- 75
13 VEIRLAQ---ISQFGVVSAPLGNMSDAALSPGKFRVAVLMVAE-----SSGGVCD 63
76 -LLDLACAVEMVHAASLIILDDMPGMDQAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134
64 AMVDAACAVEMVHAASLIIFDDMPGMDQARTTRGQPATHVAHGEGRAVLGIALITEAMRI 123
135 IARAGLTPIAKTRAVSELSTAIGMGLVQGFQKLSGDKPDSADAILLTNQFKTSTLF 194
124 IGEARGATPDQARLVAHSMRAMGVPGLCAGQDLDL---HAPDAAGIEREQDLKTGVLF 180
195 CASTOMASIAANASCEARENLRHFLSLDGLQAFQLDDLTGMDT---TGKQINQD-AGKS 250
181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFQSYDDLLDVIGDKASTGKTARDTAAPG 240

251 TLVNLGSGAVERLRQHLRLASEHLSAACQNGHSTTOLF 290
241 PKGLMAVGQMGD-VAQHYRASRAQLDELNR-----TRLF 274

RESULT 15
US-09-920-923B-2
Sequence 2, Application US/09920923B
Patent No. 6677134
GENERAL INFORMATION:
APPLICANT: Pasamontes, Luis
APPLICANT: Tsygankov, Yuri
TITLE OF INVENTION: Fermentative Carotenoid Production
FILE REFERENCE: 15464 US (C38435/125944)
CURRENT APPLICATION NUMBER: US/09/920,923B
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 08/980,832
PRIOR FILING DATE: 1997-12-01
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 295
TYPE: PRT
ORGANISM: Flavobacterium sp. R1534
US-09-920-923B-2

Query Match 27.9%; Score 425; DB 4; Length 295;
Best Local Similarity 39.6%; Pred. No. 1.8e-40;
Matches 111; Conservative 37; Mismatches 102; Indels 30; Gaps 9;
23 IDSRDLQLLPVQGERDCV-----GAAMREGTLAPKRIKIRPMLLLLTARDLGCALSHGG--- 75
13 VEIRLAQ---ISQFGVVSAPLGNMSDAALSPGKFRVAVLMVAE-----SSGGVCD 63
76 -LLDLACAVEMVHAASLIILDDMPGMDQAQMRGRPTHTQYGEHVAILAAVALLSKAFGV 134
64 AMVDAACAVEMVHAASLIIFDDMPGMDQARTTRGQPATHVAHGEGRAVLGIALITEAMRI 123
135 IARAGLTPIAKTRAVSELSTAIGMGLVQGFQKLSGDKPDSADAILLTNQFKTSTLF 194
124 IGEARGATPDQARLVAHSMRAMGVPGLCAGQDLDL---HAPDAAGIEREQDLKTGVLF 180
195 CASTOMASIAANASCEARENLRHFLSLDGLQAFQLDDLTGMDT---TGKQINQD-AGKS 250
181 VAGLEMLSIKGLDKAETEQLMAFGRLGRVFQSYDDLLDVIGDKASTGKTARDTAAPG 240
251 TLVNLGSGAVERLRQHLRLASEHLSAACQNGHSTTOLF 290
241 PKGLMAVGQMGD-VAQHYRASRAQLDELNR-----TRLF 274

Search completed: February 29, 2004, 14:54:55
Job time : 11.2076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:26:38 ; Search time 35.8246 Seconds
(without alignments)
2389.754 Million cell updates/sec

itle: US-09-941-947A-26

effect score: 1526

equences: 1 LTVCAKHVHLTGISAQQL.....HSTQLFIQAWFDKLAAYS 303

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : A Genesep29Jan04:*

- 1: genesep1980s:*
- 2: genesep1980s:*
- 3: genesep2000s:*
- 4: genesep2001s:*
- 5: genesep2002s:*
- 6: genesep2003as:*
- 7: genesep2003bs:*
- 8: genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1526	100.0	303	5	Aae22311 Pantoea s
2	1526	100.0	303	6	Abp96685 Pantoea s
3	1518	99.5	303	6	Aa016020 Pantoea s
4	1344.5	88.1	302	2	Aar07463 Polypepti
5	1344.5	88.1	302	2	Aaw82255 C. utilis
6	1344.5	88.1	302	2	Aaw87888 Protein e
7	1344.5	88.1	302	2	Aaw99098 Erwinia u
8	763.5	50.0	298	2	Aaw01120 Recombina
9	763.5	50.0	298	2	Aaw00171 N termina
10	763.5	50.0	298	2	Aaw00342 N-termina
11	763.5	50.0	298	2	Aaw32470 Erwinia h
12	763.5	50.0	307	2	Aaw01119 Geranyige
13	763.5	50.0	307	2	Aaw00170 Geranyige
14	763.5	50.0	307	2	Aaw00341 Wild type
15	763.5	50.0	307	2	Aaw32469 Geranyige
16	762.5	50.0	298	2	Aar13982 Geranyige
17	761.5	49.9	307	2	Aar13981 Geranyige
18	643	42.1	332	6	Abm70119 Photorhab
19	427	28.0	295	6	Abu97246 Enzyme po
20	425	27.9	295	2	Aaw06515 Flavobact
21	425	27.9	295	2	Aaw69530 Flavobact
22	425	27.9	295	7	ADB74152 Flavobact
23	388	25.4	369	5	Aae25932 Soybean G
24	388	25.4	369	6	Abu67329 Soybean G
25	377.5	24.7	347	5	Abb91794 Herbicida

26	376.5	24.7	287	4	AAU00859	Aau00859 S. aureus
27	376.5	24.7	293	6	ABM71047	Abm71047 Staphyloc
28	376.5	24.7	357	5	ABB92479	Abb92479 Herbicida
29	375.5	24.6	293	2	Aaw69537	Aaw69537 Flavobact
30	375.5	24.6	293	4	AAU36990	Aau36990 Staphyloc
31	375.5	24.6	293	6	ABU16439	Abu16439 Protein e
32	369	24.2	295	3	AAG42844	Aag42844 Arabidops
33	369	24.2	371	3	AAG42843	Aag42843 Arabidops
34	369	24.2	371	5	ABP99379	Abp99379 Arabidops
35	369	24.2	371	5	ABB93265	Abb93265 Herbicida
36	368.5	24.1	288	4	AAU33743	Aau33743 Staphyloc
37	368	24.1	360	5	ABB92493	Abb92493 Herbicida
38	367.5	24.1	293	3	AAE23333	Aae23333 Staphyloc
39	364	23.9	287	3	AAG07001	Aag07001 Arabidops
40	364	23.9	287	3	AAG42713	Aag42713 Arabidops
41	364	23.9	288	3	AAG42712	Aag42712 Arabidops
42	364	23.9	288	3	AAG07000	Aag07000 Arabidops
43	364	23.9	376	3	AAG42711	Aag42711 Arabidops
44	364	23.9	376	3	AAG06999	Aag06999 Arabidops
45	364	23.9	376	5	ABB91864	Abb91864 Herbicida

ALIGNMENTS

RESULT 1

AAE22311

ID AAE22311 standard; protein; 303 AA.

XX

AC AAE22311;

XX

DT 25-JUL-2002 (first entry)

XX

DE Pantoea stewartii geranylgeranyl pyrophosphate synthase (CrTE) enzyme.

XX

KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; geranylgeranyl pyrophosphate synthase; CrTE.

XX

OS Pantoea stewartii.

XX

FN WO200218617-A2.

XX

PD 07-MAR-2002.

XX

PF 04-SEP-2001; 2001WO-US027420.

XX

PR 01-SEP-2000; 2000US-0229858P.

XX

PR 01-SEP-2000; 2000US-0229997P.

XX

(DUPO) DU PONT DE NEMOURS & CO E I.

XX

Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX

Odum JM, Picataggio SK, Rouviere PE;

XX

WPI; 2002-351711/38.

XX

N-PSDE; AAD35509.

XX

Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by

XX

using microorganisms having a nucleic acid molecule encoding enzymes in

XX

the carotenoid biosynthetic pathway and which metabolize single carbon

XX

substrates.

XX

Claim 15; Page 134-135; 156pp; English.

XX

The invention relates to a method for producing carotenoid compounds. The method comprises a transformed metabolising host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the control of regulatory sequences, and contacting the host cell with carbon substrate to produce a carotenoid compound. The method is useful for producing carotenoid compounds such as antheraxanthin and astaxanthin, by using microorganism having a nucleic acid molecule encoding enzymes in

CC the carotenoid biosynthetic pathway and which metabolise single carbon
 CC substrates. The carotenoids have potent anti-oxidant properties useful in
 CC diet, and aquaculture elements. The carotenoids are also useful as
 CC intermediates in the synthesis of steroids flavours and fragrances and
 CC compounds for potential electro-optic applications. The present sequence
 CC is Pantoea stewartii geranylgeranyl pyrophosphate synthase (CrtE) enzyme
 CC used in the invention
 XX
 SQ Sequence 303 AA;

Query Match 100.0%; Score 1526; DB 5; Length 303;
 Best Local Similarity 100.0%; Pred. No. 2.6e-152;
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 2Y 1 LTVCCKKHVHTGTISABQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 DB 1 LTVCCKKHVHTGTISABQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 2Y 61 LTTARDLGCALSHGGLLDLACAVEMVHAASILLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 DB 61 LTTARDLGCALSHGGLLDLACAVEMVHAASILLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 2Y 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIQMGVLVQGFQKDLSEGDKPRSD 180
 DB 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIQMGVLVQGFQKDLSEGDKPRSD 180
 2Y 181 AILLTNQKSTLFCASTOMASIAANASCEARENLRHPSLDLGOAFOLLDDLTGMDTDTG 240
 DB 181 AILLTNQKSTLFCASTOMASIAANASCEARENLRHPSLDLGOAFOLLDDLTGMDTDTG 240
 2Y 241 KDINQACKSTLVNLLGSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIQAQFDDKLA 300
 DB 241 KDINQACKSTLVNLLGSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIQAQFDDKLA 300
 2Y 301 AVS 303
 DB 301 AVS 303

RESULT 2

ABP96685
 ID ABP96685 standard; protein; 303 AA.

XX AC ABP96685;

XX DT 03-JUN-2003 (first entry)

XX DE Pantoea stewartii geranylgeranyl pyrophosphate synthase SEQ ID NO:2.

XX KW Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;
 KW crtI; crtB; crtZ; geranylgeranyl pyrophosphate synthase; enzyme;
 KW phytoene; carotenoid.

XX OS Pantoea stewartii.

XX PN WO2003016503-A2.

XX PD 27-FEB-2003.

XX PF 15-AUG-2002; 2002WO-US026647.

XX PR 15-AUG-2001; 2001US-0312646P.

XX PA (DUFO) DU PONT DE NEMOURS & CO E. I.

XX PI Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

XX DR WPI; 2003-268323/26.

XX JR N-PSDB; ACC44759.

XX Novel nucleic acid molecule isolated from Pantoea stewartii encoding a
 PT carotenoid biosynthetic enzyme, useful for regulating carotenoid
 PT biosynthesis in an organism.

XX PS Claim 4; Page 57-58; 68pp; English.

XX The present invention describes Pantoea stewartii carotenoid biosynthetic
 CC enzymes (I). More specifically described are the geranylgeranyl
 CC pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),
 CC lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
 CC (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to
 CC ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating
 CC carotenoid biosynthesis in an organism, by over-expressing (I) in an
 CC organism, such that the carotenoid biosynthesis is altered in the
 CC organism. (II) and the genes encoding (I) are useful for converting
 CC phytoene to the carotenoids, for creating recombinant organisms that have
 CC the ability to produce various carotenoid compounds, and also for
 CC enhancing or manipulating carotenoid compounds. (I) can also be used for
 CC producing gene products having enhanced or altered activity

XX SQ Sequence 303 AA;

Query Match 100.0%; Score 1526; DB 6; Length 303;
 Best Local Similarity 100.0%; Pred. No. 2.6e-152;
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LTVCCKKHVHTGTISABQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 DB 1 LTVCCKKHVHTGTISABQLLADISRLDQLLPVQGERDCVGAAMREGTLAPGKRIRPMLL 60
 QY 61 LTTARDLGCALSHGGLLDLACAVEMVHAASILLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 DB 61 LTTARDLGCALSHGGLLDLACAVEMVHAASILLDDMPDMDAQMRRGRPTIHTQYGEHVA 120
 QY 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIQMGVLVQGFQKDLSEGDKPRSD 180
 DB 121 ILAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIQMGVLVQGFQKDLSEGDKPRSD 180
 QY 181 AILLTNQKSTLFCASTOMASIAANASCEARENLRHPSLDLGOAFOLLDDLTGMDTDTG 240
 DB 181 AILLTNQKSTLFCASTOMASIAANASCEARENLRHPSLDLGOAFOLLDDLTGMDTDTG 240
 QY 241 KDINQACKSTLVNLLGSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIQAQFDDKLA 300
 DB 241 KDINQACKSTLVNLLGSGAVEERLRQHLRLASEHLSAAQNGHSTTQLFQIQAQFDDKLA 300
 QY 301 AVS 303
 DB 301 AVS 303

RESULT 3

AAO16020
 ID AAO16020 standard; protein; 303 AA.

XX AC AAO16020;

XX DT 20-FEB-2003 (first entry)

XX DE Pantoea stewartii geranylgeranyl pyrophosphate synthase.

XX KW Carotenoid; crt.

XX OS Pantoea stewartii.

XX PN WO200279395-A2.

XX PD 10-OCT-2002.

XX PF 25-JAN-2002; 2002WO-US002124.

XX PR 26-JAN-2001; 2001US-0264329P.

XX PR 04-MAY-2001; 2001US-0288984P.

XX PA (CRGI) CARGILL INC.

De Souza ML, Kollmann SR, May CA, Schroeder WA;
WPI: 2003-075455/07.
N-PSDB; ABT14192.

Novel isolated nucleic acid useful e.g. to engineer host cells with the ability to produce particular carotenoids and polypeptides useful in cell-free systems to make particular carotenoids.

Claim 20; Page 61-62; 74pp; English.

The invention comprises the amino acid and coding sequence of a number of carotenoid (crt)-related proteins. The crt-related DNA and protein sequences of the invention are useful for engineering cells which are able to produce carotenoids. The present amino acid sequence represents a crt-related protein of the invention

Sequence 303 AA;

```
Query Match          99.5%; Score 1518; DB 6; Length 303;
Best Local Similarity 99.3%; Pred. No. 1.8e-151;
Matches 301; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
1  LTVCAKGVHVTGISAEQLADISRLDQLLPVQGERDCVGAAMREGTLPAGKRIKIRPMLL 60
1  MMVCAKGVHVTGISAEQLADISRLDQLLPVQGERDCVGAAMREGTLPAGKRIKIRPMLL 60
61  LITARDLGAISHGGLLDLACAVEMVHAASLILODMPQMDDAQMERGRPTIHTQYGEHVA 120
61  LITARDLGAISHGGLLDLACAVEMVHAASLILODMPQMDDAQMERGRPTIHTQYGEHVA 120
121  ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRASD 180
121  ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRASD 180
181  AILLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAQFOLLDDLTGMDTGTG 240
181  AILLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAQFOLLDDLTGMDTGTG 240
241  KDINDAGKSTLVNLGSGAVEERLRHRLASEHLSAACQNGHSTTQLTQAFWFKKLA 300
241  KDINDAGKSTLVNLGSGAVEERLRHRLASEHLSAACQNGHSTTQLTQAFWFKKLA 300
301  AVS 303
301  AVS 303
```

```
RESULT 4
AA07463
AA07463 standard; protein; 302 AA.
AA07463;
24-OCT-2003 (revised)
28-JAN-1991 (first entry)
X Polypeptide with enzymatic activity for the conversion of prephytoene
X Pyrophosphate into phytoene.
X Carotenoid biosynthesis; vitamin A; cancer; food coloring.
X Pantoea ananatis.
X EP393690-A.
X 24-OCT-1990.
X 20-APR-1990; 90EP-00107493.
X 21-APR-1989; 89JP-00103078.
X 05-MAR-1990; 90JP-00053255.
```

(KIRI) KIRIN BEER KK.
Misawa N, Kobayashi K, Nakamura K;
WPI: 1990-322212/43.
N-PSDB; AAQ06293.

DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.

Claim 1; Fig 1; 40pp; English.

Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

Sequence 302 AA;

```
Query Match          88.1%; Score 1344.5; DB 2; Length 302;
Best Local Similarity 88.4%; Pred. No. 3.9e-133; Indels 1; Gaps 1;
Matches 268; Conservative 16; Mismatches 18;
1  LTVCAKGVHVTGISAEQLADISRLDQLLPVQGERDCVGAAMREGTLPAGKRIKIRPMLL 60
1  MTVCAKGVHVTGISAEQLADISRLDQLLPVQGERDCVGAAMREGTLPAGKRIKIRPMLL 60
61  LITARDLGAISHGGLLDLACAVEMVHAASLILODMPQMDDAQMERGRPTIHTQYGEHVA 120
61  LITARDLGAISHGGLLDLACAVEMVHAASLILODMPQMDDAQMERGRPTIHTQYGEHVA 120
121  ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRASD 180
121  ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSTAIGMQLVQGFQKDLSEGDKPRAS 180
181  AILLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAQFOLLDDLTGMDTGTG 240
181  AILLTNQKSTLPCASTOMASIAANASCEARENLRHPSLDLGAQFOLLDDLTGMDTGTG 240
241  KDINDAGKSTLVNLGSGAVEERLRHRLASEHLSAACQNGHSTTQLTQAFWFKKLA 300
241  KDINDAGKSTLVNLGSGAVEERLRHRLASEHLSAACQNGHSTTQLTQAFWFKKLA 299
301  AVS 303
300  AVS 302
```

```
RESULT 5
AAW82255
AAW82255 standard; protein; 302 AA.
AAW82255;
17-OCT-2003 (revised)
16-JUL-1999 (first entry)
C. utilis crtE protein.
HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtE; carotenoid.
Pichia jadinii.
JP10248575-A.
22-SEP-1998.
12-MAR-1997; 97JP-00058012.
12-MAR-1997; 97JP-00058012.
(KIRI ) KIRIN BREWERY KK.
```

R WPI; 1998-560727/48.
R N-PSDB; AAV73179.
X Gene useful for increase in carotenoid production - and preparation of
T carotenoid.
X
X
X Example 2; Fig 5-7; 54pp; Japanese.
X
C This invention describes a novel method for the preparation of
C carotenoids using genes and proteins isolated from *Candida utilis*. The
C invention specifically describes the isolation of a 3-hydroxy-3-
C methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence
C represents the *Candida utilis* crtE protein which is used in the method of
C the invention. (Updated on 17-OCT-2003 to standardise OS field)
X
X
Q Sequence 302 AA;

Query Match 88.1%; Score 1344.5; DB 2; Length 302;
Best Local Similarity 88.4%; Pred. No. 3.9e-133;
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

Y 1 LTVCAKGVHLTGISAQQLADIDSLDQLLPVQGERDCVGAAMEGTLAPGKRIRPMLL 60
b 1 MTVCAGKGVHLTGISAQQLADIDSLDQLLPVQGERDCVGAAMEGTLAPGKRIRPMLL 60
Y 61 LTTARDLGCAVSHGGLLDLACAVEMVHAASLIIDDMPCMDAOMRGRPTIHTQYGEHVA 120
b 61 LTTARDLGCAVSHGGLLDLACAVEMVHAASLIIDDMPCMDAOMRGRPTIHTQYGEHVA 120
Y 121 ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSAIGMGLVQGFQKLSGDKPRSA 180
b 121 ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSAIGMGLVQGFQKLSGDKPRSA 180
Y 181 AILLTNQFKSTLFCASQMWASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
b 181 AILLTNQFKSTLFCASQMWASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
Y 241 KDINQDAGKSTLVNLLGSGAVFERLQHLRLASEHLSAACQNGHSTTQFIQAWFDKLLA 300
b 241 KDSNQDAGKSTLVNLLGSGAVFERLQHLRLASEHLSAACQNGHSTTQFIQAWFDKLLA 300
Y 301 AVS 303
b 300 AVS 302

RESULT 6
AAW87888 standard; protein; 302 AA.

AAW87888;
17-OCT-2003 (revised)
10-MAR-1999 (first entry)
Protein encoded by the carotenoid biosynthesis gene crtE.
Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;
crtI gene; adonixanthin-3-glucoside; astaxanthin monoglucoside;
carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;
food additive.
Pantoea ananatis.
JP10327865-A.
15-DEC-1998.
29-MAY-1997; 97JP-00140460.
29-MAY-1997; 97JP-00140460.
(KIRI) KIRIN BREWERY KK.

PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX WPI; 1999-099030/09.
DR N-PSDB; AAV84080.
XX New carotenoid glucoside(s) - used as food additives.
XX
XX
XX Disclosure; Page 17-18; 26pp; Japanese.

CC The present sequence represents a protein involved in carotenoid
CC biosynthesis. The specification describes astaxanthin diglucosides and
CC adonixanthin-3'-glucosides. The specification also describes a method for
CC the preparation of a carotenoid glycoside, in which all, or part of,
CC carotenoid biosynthesis genes crtE, crtB, crtI, crtZ, crtX or crtW
CC are introduced to a microbe or plant and expressed. The transformed
CC organism is cultured and astaxanthin diglucosides, adonixanthin-3'-
CC glucosides, and/or astaxanthin monoglucoside are collected. The
CC carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003
CC to standardise OS field)

Sequence 302 AA;

Query Match 88.1%; Score 1344.5; DB 2; Length 302;
Best Local Similarity 88.4%; Pred. No. 3.9e-133;
Matches 268; Conservative 16; Mismatches 18; Indels 1; Gaps 1;

QY 1 LTVCAKGVHLTGISAQQLADIDSLDQLLPVQGERDCVGAAMEGTLAPGKRIRPMLL 60
Db 1 MTVCAGKGVHLTGISAQQLADIDSLDQLLPVQGERDCVGAAMEGTLAPGKRIRPMLL 60
QY 61 LTTARDLGCAVSHGGLLDLACAVEMVHAASLIIDDMPCMDAOMRGRPTIHTQYGEHVA 120
Db 61 LTTARDLGCAVSHGGLLDLACAVEMVHAASLIIDDMPCMDAOMRGRPTIHTQYGEHVA 120
QY 121 ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSAIGMGLVQGFQKLSGDKPRSA 180
Db 121 ILAAVALLSKAFGVIAEAGLTPIAKTRAVSELSAIGMGLVQGFQKLSGDKPRSA 180
QY 181 AILLTNQFKSTLFCASQMWASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
Db 181 AILLTNQFKSTLFCASQMWASIVANASSEARCLHRFSLDLGQAFQLLDLTGMDTGT 240
QY 241 KDINQDAGKSTLVNLLGSGAVFERLQHLRLASEHLSAACQNGHSTTQFIQAWFDKLLA 300
Db 241 KDSNQDAGKSTLVNLLGSGAVFERLQHLRLASEHLSAACQNGHSTTQFIQAWFDKLLA 300
QY 301 AVS 303
Db 300 AVS 302

RESULT 7
AAW99098 standard; protein; 302 AA.

AAW99098;
17-OCT-2003 (revised)
14-MAY-1999 (first entry)
Erwinia uredovora crtE protein sequence.
Beta-carotene hydroxylase; crtY; crtB; crtE; crtI; xanthophyll;
metabolite.
Pantoea ananatis.
JP11046770-A.
23-FEB-1999.
07-AUG-1997; 97JP-00213648.

synthase. GGPP synthase is a enzyme which, along with phytoene synthase (see also AAW00343), catalyze the formation of carotenoids, particularly phytoene. The N-terminally truncated form of GGPP synthase has the first thirteen amino acids replaced by four heterologous amino acids derived from the plasmid PARC306A. This enzyme is about twice as active as the wild type enzyme. A C-terminal truncated protein was found to have even higher activity and was cloned into plasmid PARC489D. The GGPP synthase coding sequences may be attached to the transit peptide sequence given in AAW00344, for transport into tobacco chloroplasts. This allows the tobacco plants to produce high levels of phytoene for use in the treatment of skin disorders. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field)

Sequence 298 AA;

Query Match 50.0%; Score 763.5; DB 2; Length 298;
 Best Local Similarity 56.0%; Pred. No. 9.4e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
 17 EQLLADIDSELQDLFPVQGERDCVGAAMEGTLAPGKIRPMLLLTARDLGCAISHGGL 76
 7 EVMRQSIDDHLAGLLPETDSQDIVSLAMREGVMAPGKIRPMLLLTARDLRYQGSMPPL 66
 77 LDIACAVEMVHAAHLIDMPDMDAOMRGPTTHQYGEHVAILAAVALLSKAFGVIA 136
 67 LDIACAVELTHTASLMDDMPDMDAOMRGPTTHQYGEHVAILAAVALLSKAFGLIA 126
 137 EABGLTPIAKTRAVSELSTAIGMQGLVQGFQKDLSEGDKPRSDAILLTNPKTSTLPCA 196
 127 ATGDLPCERRAQAVNELSTAVGLQGLVGFQFRLDNDALDRTDAILSTNHLKTGLFSA 186
 197 STOMASIANASCEARENLRPSLDLQGFQFLLDLDLTDGKTTGKDNODACKSTLVNLL 256
 187 MLQIVAIASASSPSTRETLHAFALDFQAFQGLLDLDLDDHDPETGKDRNKDACKSTLVNRL 246
 257 GSGAVERLRQHLRLASEHLSAACQNGHSTTOLFQAWFDKXLA 300
 247 GADAAKQKUREHIDSADKHLTPACPOGGAIRO-FMHLWFGHILA 289

BULT 11
 #32470

AAW32470 standard; protein; 298 AA.

AAW32470;

17-OCT-2003 (revised)
 25-MAR-2003 (revised)
 15-JAN-1998 (first entry)

Erwinia herbicola geranylgeranyl pyrophosphate synthase.

Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP; lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene; yeast; plant; vitamin A; cancer.

Pantoea agglomerans.

Key Location/Qualifiers
 Misc-difference 91
 /note= "Encoded by ACC"

US5656472-A.

12-AUG-1997.

07-JUN-1995; 95US-00473512.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00526674.

28-FEB-1991; 91US-00662921.

21-JUL-1993; 93US-00095726.

XX (STAD) AMOCO CORP.
 XX PA
 XX Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB
 PI Brinkhaus FL;
 XX
 XX WPI; 1997-414592/38.
 DR N-PSDB; AAT91542.
 XX
 XX DNA encoding Erwinia herbicola lycopene cyclase - for producing
 PT recombinant enzyme, and transgenic organisms with increased beta-carotene
 PT levels.
 XX
 XX Example 2; Fig 3; 102pp; English.

XX A novel DNA molecule has been isolated which encodes an Erwinia herbicola
 CC lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA
 CC molecule comprises at least 1125 bp and is present in the plasmids
 CC PARC147, PARC1509, PARC1510 and PARC1520. The present sequence represents
 CC the amino acid sequence corresponding to the preferred heterologous
 CC structural gene of Erwinia herbicola geranylgeranyl pyrophosphate (GGPP)
 CC synthase. The new DNA molecule can be used to produce the recombinant
 CC enzyme and transgenic organisms, e.g. yeasts or plants, with increased
 CC beta-carotene levels. Beta-carotene is used as a colourant in margarine
 CC and butter and as an intermediate for vitamin A, and may prevent cancer.
 CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 17-OCT-2003 to
 CC standardise OS field)

XX Sequence 298 AA;

Query Match 50.0%; Score 763.5; DB 2; Length 298;
 Best Local Similarity 56.0%; Pred. No. 9.4e-72;
 Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;
 QY 17 EQLLADIDSELQDLFPVQGERDCVGAAMEGTLAPGKIRPMLLLTARDLGCAISHGGL 76
 Db 7 EVMRQSIDDHLAGLLPETDSQDIVSLAMREGVMAPGKIRPMLLLTARDLRYQGSMPPL 66
 QY 77 LDIACAVEMVHAAHLIDMPDMDAOMRGPTTHQYGEHVAILAAVALLSKAFGVIA 136
 Db 67 LDIACAVELTHTASLMDDMPDMDAOMRGPTTHQYGEHVAILAAVALLSKAFGLIA 126
 QY 137 EABGLTPIAKTRAVSELSTAIGMQGLVQGFQKDLSEGDKPRSDAILLTNPKTSTLPCA 196
 Db 127 ATGDLPCERRAQAVNELSTAVGLQGLVGFQFRLDNDALDRTDAILSTNHLKTGLFSA 186
 QY 197 STOMASIANASCEARENLRPSLDLQGFQFLLDLDLTDGKTTGKDNODACKSTLVNLL 256
 Db 187 MLQIVAIASASSPSTRETLHAFALDFQAFQGLLDLDLDDHDPETGKDRNKDACKSTLVNRL 246
 QY 257 GSGAVERLRQHLRLASEHLSAACQNGHSTTOLFQAWFDKXLA 300
 Db 247 GADAAKQKUREHIDSADKHLTPACPOGGAIRO-FMHLWFGHILA 289

RESULT 12

AAW01119

ID AAW01119 standard; protein; 307 AA.

XX AAW01119;

XX 16-OCT-2003 (revised)

DT 25-MAR-2003 (revised)

DT 10-DEC-1996 (first entry)

XX Geranylgeranyl pyrophosphate synthase.

XX GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;
 KW phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; d-glucoside;
 KW pigment; food colourant; chloroplast transit peptide; increase yield;
 KW tobacco ribulose bis-phosphate carboxylase-oxygenase.

XX Pantoea agglomerans.

OS

16 EVMRQSIDHLAGLIPETDSQDIVSLAWREGVAPGKIRPMLLLTARDLRYQGSMTL 75
 77 LDACAVEMVHAASLILDDMPQDDAQRGRPIHTQYGEHVAILAAVALLSKAFVIA 136
 76 LDACAVELTHTASLMDLDDMPQDDAQRGRPIHTQYGEHVAILAAVALLSKAFVIA 135
 137 EAEGLTPIAKTRAVSELSTAIGMGLVQGFQKLSSEGDKPSADAILTNQFKSTLFC 196
 136 ATGDLPGERRAQAVALNELSTAVGLQGLVGFQFRLDNDALDTPDAILSTNHLKTGLFSA 195
 197 STOMASIAANASCEARENHRSFLDLGOAFQFLLDITGMDTDTGKIDNQDAGKSTLVNLL 256
 196 MLQIVAIASASSPSTRETTHAFALDFGQAFQFLLDITGMDTDTGKIDNQDAGKSTLVNRL 255
 257 GSGAVEERLROHLRLASEHLSAACQNGHSTTQLFQIAWFDKLA 300
 256 GADAAQKRLREHIDSADKHLTFACPGGAIRQ-FMELWFGHLLA 298

RESULT 14

AAW00341 standard; protein; 307 AA.

AAW00341;

16-OCT-2003 (revised)

25-MAR-2003 (revised)

10-DEC-1996 (first entry)

Wild type geranylgeranyl pyrophosphate synthase.

Wild type; N-terminally truncated; geranylgeranyl pyrophosphate; GGPP;

GGPP synthase; phytoene synthase; carotenoid; phytoene; PARC306A;

PARC489D; transit peptide; tobacco; chloroplast; skin disorder.

Pantoea agglomerans.

US5545816-A.

13-AUG-1996.

19-JUL-1993; 93US-00093577.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

30-OCT-1991; 91US-00785569.

(STAD) AMOCO CORP.

Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;

Ausich RL;

WPI; 1996-383742/38.

N-PSDB; AAT41741.

Transformed plants containing DNA encoding Erwinia herbicola enzymes - esp. geranylgeranyl pyrophosphate synthase and phytoene synthase, allows large scale production of phytoene.

Disclosure; Fig 2; 61pp; English.

The sequences given in AAW00341-42 represent the wild type and an N-

terminally truncated form of geranylgeranyl pyrophosphate (GGPP)

synthase. GGPP synthase is a enzyme which, along with phytoene synthase

(see also AAW00343), catalyze the formation of carotenoids, particularly

phytoene. The N-terminally truncated form of GGPP synthase has the first

thirteen amino acids replaced by four heterologous amino acids derived

from the plasmid PARC306A. This enzyme is about twice as active as the

wild type enzyme. A C-terminal truncated protein was found to have even

higher activity and was cloned into plasmid PARC489D. The GGPP synthase

coding sequences may be attached to the transit peptide sequence given in
 CC AAW00344, for transport into tobacco chloroplasts. This allows the
 CC tobacco plants to produce high levels of phytoene for use in the
 CC treatment of skin disorders. (Updated on 25-MAR-2003 to correct PF
 CC field.) (Updated on 16-OCT-2003 to standardise OS field)

XX SQ Sequence 307 AA;

Query Match 50.0%; Score 763.5; DB 2; Length 307;

Sest Local Similarity 56.0%; Pred. No. 9.8e-72;

Matches 159; Conservative 40; Mismatches 84; Indels 1; Gaps 1;

QY 17 EQLADIISRLDQLLPVQGERDCVCAANREGTLAPGKIRPMLLLTARDLGAISHGGL 76

Db 16 EYMRQSIDHLAGLIPETDSQDIVSLAWREGVAPGKIRPMLLLTARDLRYQGSMTL 75

QY 77 LDACAVEMVHAASLILDDMPQDDAQRGRPIHTQYGEHVAILAAVALLSKAFVIA 136

Db 76 LDACAVELTHTASLMDLDDMPQDDAQRGRPIHTQYGEHVAILAAVALLSKAFVIA 135

QY 137 EAEGLTPIAKTRAVSELSTAIGMGLVQGFQKLSSEGDKPSADAILTNQFKSTLFC 196

Db 136 ATGDLPGERRAQAVALNELSTAVGLQGLVGFQFRLDNDALDTPDAILSTNHLKTGLFSA 195

QY 197 STOMASIAANASCEARENHRSFLDLGOAFQFLLDITGMDTDTGKIDNQDAGKSTLVNLL 256

Db 196 MLQIVAIASASSPSTRETTHAFALDFGQAFQFLLDITGMDTDTGKIDNQDAGKSTLVNRL 255

QY 257 GSGAVEERLROHLRLASEHLSAACQNGHSTTQLFQIAWFDKLA 300

Db 256 GADAAQKRLREHIDSADKHLTFACPGGAIRQ-FMELWFGHLLA 298

RESULT 15

AAW32469

ID AAW32469 standard; protein; 307 AA.

XX AC AAW32469;

XX 25-MAR-2003 (revised)

DT 15-JAN-1998 (first entry)

XX Geranylgeranyl pyrophosphate synthase.

XX Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;

XX lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;

XX yeast; plant; vitamin A; cancer.

XX Erwinia sp.

XX US5656472-A.

XX 12-AUG-1997.

XX 07-JUN-1995; 95US-00473512.

XX 02-MAR-1990; 90US-00487613.

XX 18-MAY-1990; 90US-00525551.

XX 03-AUG-1990; 90US-00562674.

XX 28-FEB-1991; 91US-00662921.

XX 21-JUL-1993; 93US-00095726.

XX (STAD) AMOCO CORP.

XX Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;

XX Brinkhaus FL;

XX WPI; 1997-414592/38.

XX N-PSDB; AAT91541.

XX DNA encoding Erwinia herbicola lycopene cyclase - for producing
 PT recombinant enzyme, and transgenic organisms with increased beta-carotene
 PT levels.

Example 2; Fig 2; 102pp; English.

A novel DNA molecule has been isolated which encodes an Erwinia herbicola lycopene cyclase enzyme that converts lycopene to beta- carotene. The DNA molecule comprises at least 1125 bp and is present in the plasmids pARC147, pARC1509, pARC1510 and pARC1520. The present sequence represents amino acid sequence corresponding to the nucleotide base sequences of certain preferred DNA segments of the structural gene for geranylgeranyl pyrophosphate (GGPP) synthase. The new DNA molecule can be used to produce the recombinant enzyme and transgenic organisms, e.g. yeasts or plants, with increased beta-carotene levels. Beta- carotene is used as a colouant in margarine and butter and as an intermediate for vitamin A, and may prevent cancer. (Updated on 25-MAR-2003 to correct PF field.)

Sequence 307 AA;

Query Match	50.0%;	Score 763.5;	DB 2;	Length 307;
Best Local Similarity	56.0%;	Pred. No. 9.8e-72;		
Matches 159;	Conservative	40;	Mismatches 84;	Indels 1;
Gaps 1;				
17	ECLLADIDSLDQLLPVQGERDCVGAMREGTLAPGKIRPMLLLLTARDLGCAISHGGL	76		
16	EVWQSIDDHLAGLLPETDSQDIVSLANREGVWAPGKIRPMLLLAARDLRYQGSMPFL	75		
77	LDLACAVEMVHAASLIEDDMPCHDDAQMRRGRPTIHTQYGEHVAILAAVALLSKAFGVIA	136		
76	LDLACAVELTHTASIMLDDMPCHMDNAELRERGQPTTHKXGCVAILASVGLLSKAFGLIA	135		
137	EAEGLTPIAKTRAVSELSAIGMGLVQGFOLSEGDKPSADAILLTNQFTSTLPCA	196		
136	ATGGLPGERRAQAVNELSTAVGLQGLVIGQFRLNDALDRTDAILSTNHLKTGLFSA	195		
197	STOMASTAANASCEARENHRFSLDLGQAFOLLDDLTGDMTDGKINODAGKSTLVNLL	256		
196	MLQIVAIASASSPSTRTLHAPALDFQAFQLLDDLDHDPETGKDRNKDAGKSTLVNEL	255		
257	GSGAVEERLRQHLSEHLSAACONGHSTTQLFIQAWFDKILA	300		
256	GADAAKQKLRHIDSADKHLTFACPGGAIKQ-FMHLWFGHILA	298		

Search completed: February 29, 2004, 14:43:51
Job time : 39.9246 secs

result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1337	51.5	494	16	Q8ENM2	Q8ENM2	oceanobacil.
2	1287	49.5	457	16	Q9R73	Q9R73	staphylococ
3	1280	49.3	437	16	Q8NUQ3	Q8NUQ3	staphylococ
4	1050	40.4	439	2	Q53589	Q53589	staphylococ
5	888.5	34.2	534	16	Q70U6E	Q70U6E	rhodopirell
6	779	30.0	517	2	Q9ZGB4	Q9ZGB4	heliobacill
7	731	28.1	511	16	Q8ENM0	Q8ENM0	oceanobacil
8	724	27.9	492	2	Q8VUJ5	Q8VUJ5	pantosa agg
9	706	27.2	504	2	Q93CI7	Q93CI7	xanthobacte
10	687	26.4	492	2	Q8GCS1	Q8GCS1	pantosa ste
11	685	26.4	492	2	Q47845	Q47845	pantosa agg
12	672	25.9	502	16	Q9R76	Q9R76	staphylococ
13	672	25.9	502	16	Q8NUQ6	Q8NUQ6	staphylococ
14	667	25.7	506	2	Q07855	Q07855	staphylococ
15	662.5	25.5	512	17	Q9KPD8	Q9KPD8	halobacteri
16	658.5	25.3	506	2	Q9KIX2	Q9KIX2	bradyrhizob

QY 127 QRFLDYSKNLCTETETAGYFAKGLDGFWDLLKFGYGPLRSLLSF

OS Oceanobacillus ineyensis.
OC Bacteria: Firmicutes: Bacillales: Bacillaceae: Oceanobacillus.

122 KSFLQYAKQIDDDITTKYFDPKGLDITWBIHQSHQMLQSLKKFPLDTSWYEGIERITNPK 181
187 LVEILNFIKYVGSSPDAPALNNLLPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 246
182 LRMLAYFIKYVGSSPDAPALNNLLPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 241
247 RLDAEYSEIQKQD-GRACAVKLANGDVLPAIYVSNMEVIPAMEKILRSFASSELKKM-QR 304
242 HTNSEVTKLKKDSTGNVIAATLADDSSEIKGDFISNMEVIPYKELMEKSYIKLTKK 301
305 FPCSCGLVHLGVDRLYQOLAHNFFYSDHPRHFDVAFKSHRLSDPTTYLVAPCKTD 364
302 YEFSSGGLVHLGVDRLYQOLAHNFFYSDHPRHFDVAFKSHRLSDPTTYLVAPCKTD 361
365 PAQAPA-GCEIILKILPHIPLDPLKLLTAEDYSALRERVLVLERMGLTDLRQHVITSEY 423
362 PNQVPGGVENIKILPHIPIQ-DKFFSDDDYKQFREQVLKLENKMGHGLRESIVTDEM 420
424 WPLDIOAKYYSNQSIGYGVADRFKVLGFKAPQSSSELNLYFVGGSWNPGGMPVTL 483
421 WTPNDQSTYYSKSIYGTLSNKKINRGFKSPKSSKYDNLFFVGGSWNPGGMPVTL 480
484 SGQLVRDKI 492
481 SGQGVSEKI 489

RESULT 2

Q99R73 ID Q99R73 PRELIMINARY; PRT; 497 AA.
AC Q99R73
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SAV2564.
GN SAV2564 OR SA2351.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879,
RN (1)
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=2131952; PubMed=1418446;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayaashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240 (2001).
DR EMBL; AP003365; BAB58726.1; -;
DR EMBL; AP003137; BAB43655.1; -;
DR PIR; E90061; E90061.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phytin dehydro; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 497 AA; 57174 MW; E20EB9DDF5141C9D CRC64;

Query Match 49.5%; Score 1287; DB 16; Length 497;
Best Local Similarity 51.9%; Pred. No. 5.1e-93;
Matches 254; Conservative 94; Mismatches 143; Indels 2; Gaps 2;
QY 7 QRVTIVAGLGLSAAISLATAGFSVQLIEKNDKVGKLNITWKDGTFFDLGPSILTMHP 66
DB 3 KHIIVGGGLGGISAAIRMAQSGYSVLSYEQNTHIGKVNRESHSDGFGFDLGPSTLTPY 62

QY 67 IFEALFTGAGNMADYVQIQKVEPHRNFFEDGSVIDLCEDAETQRRELDKLGCGTYAQF 126
DB 63 IFEKLFEYSKQMSDYVITIKELPHQWLSFPDGTITDLVEGIKETQCHNAILSKQDIEEL 122
QY 127 QRFLDYSKNLCTETAGYFARGDGFMDLLKFKPLRSLLSFDVFRSMQGVRRFISDEPK 186
DB 123 QMYLNTSRIDRIETEGYVNYGLDITLSLIIKFKHPLNALINYDVHTMQQAIDKRISNPY 182
QY 187 LVEILNFIKYVGSSPDAPALNNLLPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 246
DB 183 LRQMLGYFIKYVGSSPDAPALNNLLPYQYHYGLWYKGYGMAQAMEKLAVELGVEI 242
QY 247 RLDAEYSEIQKQD-GRACAVKLANGDVLPAIYVSNMEVIPAMEKILRSFASSELKKM-QR 305
DB 243 HTGARVDNIKTYQRRVGTGRLDTGSEFVKADYIISNMEVITYKVLITLDTQRLNKLREF 302
QY 306 BPSCGLVHLGVDRLYQOLAHNFFYSDHPRHFDVAFKSHRLSDPTTYLVAPCKTD 365
DB 303 BPSSGGLVHLGVDRLYQOLAHNFFYSDHPRHFDVAFKSHRLSDPTTYLVAPCKTD 362
QY 366 AQAPAGCEIILKILPHIPLDPLKLLTAEDYSALRERVLVLERMGLTDLRQHVITSEY 425
DB 363 TCAPVGVENIKILPHIPIQ-DQPTTDEYAKFRDKILDKLEKMLTDLRKHIIYEDVMT 421
QY 426 PLDIOAKYYSNQSIGYGVADRFKVLGFKAPQSSSELNLYFVGGSWNPGGMPVTL 485
DB 422 FDEIKYRNRGAIYGVADRFKVLGFKAPQSSSELNLYFVGGSWNPGGMPVTL 481
QY 486 QLVRDKIVA 494
DB 482 QGVADKINA 490

RESULT 3

Q8NUQ3 ID Q8NUQ3 PRELIMINARY; PRT; 497 AA.
AC Q8NUQ3
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MW2485 protein.
GN MW2485
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamanoto K., Hiramatsu K.,
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827 (2002).
DR EMBL; AP004830; BAB96350.1; -;
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR ProDom; PD139017; Phytin dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 497 AA; 57217 MW; 70814ED59449BBFF CRC64;

Query Match 49.3%; Score 1280; DB 16; Length 497;
Best Local Similarity 51.5%; Pred. No. 1.8e-92;
Matches 252; Conservative 86; Mismatches 149; Indels 2; Gaps 2;
QY 7 QRVTIVAGLGLSAAISLATAGFSVQLIEKNDKVGKLNITWKDGTFFDLGPSILTMHP 66
DB 3 KHIIVGGGLGGISAAIRMAQSGYSVLSYEQNTHIGKVNRESHSDGFGFDLGPSTLTPY 62
QY 67 IFEALFTGAGNMADYVQIQKVEPHRNFFEDGSVIDLCEDAETQRRELDKLGCGTYAQF 126


```
QY 346 SHRLSDDPTIYVAPCKTDPAPAGCBIIKILPHIHLDPDKLLTAEDYSAL-----RER 401
Db 378 KCQAPADPTAVCAPAISPPVAPDGCSEIIVLWTFYLRP-----GHDWKMLPGVRDV 432
QY 402 VLVKLER-MGUTDLQHOHIVTEYTFPLDIOAKYNSNGSIYGVVADRFKNGKAPORSS 460
Db 433 ILDKLERTAGMEGLDAIVTSDLSLTPEGIHNRVYVLAAGIYGLASHGKFTCAFKPGNRK 492
QY 461 ELSNLYFVGGSNPGGMPTLSCQVLRDKIVADL 496
Db 493 DLHGLYLAGGAHPGPGMPLMSGNIAUSDQDV 528

RESULT 6
Q9ZGE4 PRELIMINARY; PRT; 517 AA.
ID Q9ZGE4
AC Q9ZGE4
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Diapophytoene dehydrogenase CrtN.
GN CrtN.
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
CX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99061957; PubMed=9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
a major photosynthesis gene cluster from Helicobacillus mobilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
DR EMBL; AF080002; AAC84034.1; -.
DR PIR; T31463; T31463.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:006118; P:electron transport; IEA.
DR InterPro; IPR000759; Adrxn_reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008151; Phytin_dehydro.
DR InterPro; IPR001100; Pyr_redox.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00411; PNDRDTASEI.
DR ProDom; PDL39017; Phytin_dehydro; 1.
DR SEQUENCE 517 AA; 58340 MW; BB933977999C4587 CRC64;

Query Match 30.0%; Score 779; DB 2; Length 517;
Best Local Similarity 34.7%; Pred. No. 7.2e-53;
Matches 174; Conservative 107; Mismatches 196; Indels 24; Gaps 8;

QY 8 RVIVTGAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDGFDFLGPSTLTWPHI 67
Db 10 RVIVGSGAGGSAAVRLANQGVDTVLEKATPGRLSALQAGSIDVCGPTIMNDV 69
QY 68 FEALFTGAGKNADYVOIQKVEPHRNPFEDSGSIDLCEDAETQRRELDKLGPGT----- 122
Db 70 FHQYFKDLGSENIEDYLDVRVNPVCHLHFTDGTCKPSIDLKELDLRISFNPDDVDGYL 129
QY 123 --YAFQFPLDYSKNKLCTETAGYFAKGLDGFWDLLKPYGLRSLLSDFDPSRDXQVRR 180
Db 130 RYLAQIHRRYQVAREKPFIEKS---FTKPSD-PFNIDTLIG---MMQLRTLNKYYDDIAR 181
QY 181 FTSDPKPLVILNYFIKYVGSSPYDAPALNNLIPYIOYH-GLWVYKGMYGMAQAMEKLA 239
Db 182 FIKDELRLSLTFQAILVGVSPFDAPSIVTLIGVYEHLSGVWYPKGMNAITQALVKLL 241
QY 240 VELGWVIRLDAVSEIQKQDGRACAVKLANGDVLPADIYVNSMEVIPAMEKLLRSPASEL 299
Db 242 GFEGGSLRVNAEVEQILIEQGGSAVGRVLANGVELKADYVVISNADFPYTMENLV--PASHR 299
```

```
QY 300 -----KMGREFEPCSGLVHLGLVDRLYQLAHNNFFVSDHPRHFDVAFKSHRLSDDPT 354
Db 300 GKRYTQKLSMEVTCGAPMLYGVNRRYDNLRLHNIYFTPDYKYSMDLEFTRQQLPQDPA 359
QY 355 IYLVAPCKTDPAPAGCBIIKILPHIHLDPDKLLTAEDYSALRERVLVLERMGLTDL 414
Db 360 MYVREPTKYDOSVAPPGKIIYLVLPVNL-SGIDWKKETHRYRELVIKLERQGVTDL 418
QY 415 RGHIVTEBYTFLDIOAKYNSNGSIYGVVADRFKNGKAPORSSLSNLYFVGGSNVP 474
Db 419 SKHIEFERIYTPETQNRFNNTYQGAAPLAPSLFQSGVFRPHIKSKEVNPVLFSGASVHP 478
QY 475 GCGMPVYTLSCQVLRDKIVAD 495
Db 479 GCGVPVVLVCGKLVSQVLAD 499

RESULT 7
Q8ENM0 PRELIMINARY; PRT; 511 AA.
ID Q8ENM0
AC Q8ENM0
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase (phytoene desaturase) (EC 1.3.-).
GN OB2461.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
CX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
Ridge and its unexpected adaptive capabilities to extreme
environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004601; BAC14417.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:006118; P:electron transport; IEA.
DR InterPro; IPR000759; Adrxn_reductase.
DR InterPro; IPR002937; Amino oxidase.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR Oxidoreductase; Complete proteome.
DR SEQUENCE 511 AA; 58463 MW; 8E70FD9F71D406AB CRC64;

Query Match 28.1%; Score 731; DB 16; Length 511;
Best Local Similarity 32.9%; Pred. No. 4.4e-49;
Matches 174; Conservative 122; Mismatches 161; Indels 72; Gaps 18;

QY 6 NRVIVTGAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDGFDFLGPSTLTWPHI 65
Db 3 NKKVIVGAGVAGLASALRLQHAGENVEVEKESMPGSGHQITKQGFDFLGPSTVMP 62
QY 66 HIFEALFTGAGKNADYVOIQKVEPHRNPF-----ED-----GSVIDLCEDAETQRRELD 116
Db 63 ELYREVVELTGRNPDYIPMERIDPMYVNPFGNTPEDRYQISLDLIQDIEA-ISDKD 121
QY 117 KLGPGTYAQ--PQRLDYSKNKLCTETAGYFAKGLDGF-----WDLKFGYP----- 161
Db 122 MAGFTYLNKLYQRF-----QIAK-DHIQRFPHPW---QFTPTKLTWN 162
QY 162 ---LRSLLSFVFRSMOQGVRRFISDPKLVILNYFIKYVGSSPYDAPALNNLIPYIOYH 218
Db 163 GLKRLATLGNADNF-----INKYINDESLRQMSFQTLIYIGISPYNGSLYTIIPMISL 216
QY 219 YGLWVYKGMYGMAQAMEKLAVELGVIRLDAVSEIQKQDGRACAVKLANGDVLPADIY 278
Db 217 YGVNVIKGMNYMATSMKLFELGGLTHYNSPVKVIIEKKATGKLENLEI-NADYV 275
```

```

279 VSNMEVIPAMEKILRSPASELK----KMQRPEPSCGLVHLGVDRLYPOLAH-HNFFYS 333
276 VCNADFPYSMENLIQDIKAKGKYTDKIDKMKYSCSCFIMVLGANKKYDEVSNVNFPS 335
334 DHPREHFAVFKSHRLSDDPPIYLVAPCKTDPAPAGCEIILPHILPHIDEDKLLTA- 392
336 ENLKETMDIDFTGREL-EDPAFYTVIGSKMDPSLAPECKDGIYVL--VPVSD---LSTAQ 389
393 -----EDYSALRERVLKLRMG-LTDLROKHIVTEYWTPLDIOAKYYSNQSGSYGVVAD 446
390 YSWGDETAYTREKVFEXKLDEIDFKGIDKDEIVTWTPLDFESKNAYNGACFGLRPT 449
447 RPNLGFAPQORSSLSNLYFVGGSVNPGGMMVTLGSLQVLRDKIVAD 495
450 LQSNHLRPOSKAKCNENLYTGSSTHGPAGVPIVLLSAXIATBELKLD 498

```

RESULT 8
 ID Q8VUJ5 PRELIMINARY; PRT; 492 AA.
 DT 01-MAR-2002 (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 GN CrtI protein.
 OS Pantoea agglomerans pv. milletiae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 NCBI_TaxID=182454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Kaminunten H., Hirata R.;
 RL "Isolation and characterization of carotenoid biosynthesis genes from
 Pantoea agglomerans pv. milletiae wist 801";
 Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 EMBL; AB076662; BAB79603.1;
 GO; GO:0016491; P:oxidoreductase activity; IEA.
 GO; GO:0003058; P:biocatalysis; IEA.
 GO; GO:0008150; B:acetyl-CoA catabolic process; IEA.
 InterPro; IPR002937; Amino oxidase.
 InterPro; IPR008151; B:acyl-CoA catabolic process.
 Pfam; PF01593; Amino oxidase; 1.
 ProDom; PD139017; Phytin dehydro; 1.
 PROSITE; PS00982; PHYTOENE_DH; 1.
 R SEQUENCE 492 AA; 54802 MW; C9BE7DB0D96E31095 CRC64;

```

Query Match 27.9%; Score 724; DB 2; Length 492;
Best Local Similarity 33.5%; Pred. No. 1.5e-48;
Matches 166; Conservative 93; Mismatches 224; Indels 12; Gaps 5;

```

Y 8 RVIVIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDGTFTDLGSPILTWHPI 67
 b 3 RTTVIGAGFGGLAALRLQASGVPTLLEQDKPGGRAYVYQDQGTFTDAGTFTVITDPSA 62
 Y 68 RFALFTGACKNMADYVQIKVPEPHRNPFDGSGVIDLCEAETORRELDKLGPGTYAQF 127
 b 63 IEELFTLAGKLSIDYVELNPNVFPFRLCWESKVSFNDQPALEAQAFAFNPRDVEGYR 122
 Y 128 RFLDYSKNLCTETETAGYFAKGLDGDWLLKFGYPLRS---LLSPVFRSMQGVRRFISD 184
 b 123 RFLAYSEAVFAE---GYLKLGTVPF---LSFEDMLRAAPQLAKLQAMRSVYSKVASIED 176
 Y 185 PKLVEILNYFIKVGSSPYDAPALMNLPIYIOHYGLAVVKGMYGMAQAMEKLAVELGV 244
 b 177 EHLRQAFSFLVLVGNPFATSIYTLIHALERNGWVFPFGGTGALVQGMVKLFEDLGG 236
 Y 245 EIRLDAEVESEIQKQGRACAVKLANGDVLADIVVSNMEVIPAMEKIL-RSPASEL--KK 301
 b 237 EVELNASVARLEQENRIATVHLKDGKRVFPPTRAVNASADVHVHTYRELLSQHPASQAGRS 296
 Y 302 MQRPEPSCGLVHLGVDRLYPOLAHNFFYSDHPRHFDAVFKSHRLSDDPPIYLVAPC 361

```

297 LQKEMSNLSFVIYFELNHHNQLAHHTVCFPRYRELIDEIFNKDGLAEDFSLYLHAPC 356
362 KTDPAQAPAGCEIILPHILPHIDEDKLLTAEDYSALRERVLKLRMGGLTDLRQHIYTE 421
357 VTDPSLAPEGSGSYVLAAPVPHLGTADIDNAVEGPRLARDRIFDYLEQHYMPGLRSQLVTH 416
422 EYWTPLDIOAKYYSNQSGSYGVVADRFKNLGFAPQORSSLSNLYFVGGSVNPGGMMV 481
417 RIPTFTDFRDELNAVQGSFAFSEPIITQSAMFRPHNRDNINNLVYGAGTHPGAGIPGV 476
482 TLSSGLVLRDKIVADL 496
477 IGSAKATAGLMLDLDL 491

```

RESULT 9
 ID Q93CI7 PRELIMINARY; PRT; 504 AA.
 DT 01-DEC-2001 (TremBLrel. 19, Created)
 DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE Phytoene dehydrogenase.
 GN CRTI.
 OS Xanthobacter sp. (strain Pv2).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhodospirillaceae; Xanthobacter.
 NCBI_TaxID=78245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pv2;
 RA Larsen R.A., Metcalf W.W.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF408848; AAL02000.1;
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0003058; P:biocatalysis; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 InterPro; IPR002937; Amino oxidase.
 InterPro; IPR008150; B:acetyl-CoA catabolic process.
 InterPro; IPR008151; B:acyl-CoA catabolic process.
 Pfam; PF01593; Amino oxidase; 1.
 ProDom; PD139017; Phytin dehydro; 1.
 DR PROSITE; PS00982; PHYTOENE_DH; 1.
 R SEQUENCE 504 AA; 55556 MW; CA848D11A6C7A828 CRC64;

```

Query Match 27.2%; Score 706; DB 2; Length 504;
Best Local Similarity 33.1%; Pred. No. 4.1e-47;
Matches 163; Conservative 87; Mismatches 237; Indels 6; Gaps 2;

```

Y 7 QRIVIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLNIMTKDGTFTDLGSPILTWHPI 66
 b 10 RRAAVIGSGFGGLAALRLQASGVPTLLEQDKPGGRAYVYQDQGTFTDAGTFTVITDPT 69
 Y 67 RFALFTGACKNMADYVQIKVPEPHRNPFDGSGVIDLCEAETORRELDKLGPGTYAQF 126
 b 70 CLEEVFAAAGKLSIDYVELNPNVFPFRLCWESKVSFNDQPALEAQAFAFNPRDVEGYR 129
 Y 127 QRFLDYSKNLCTETETAGYFAKGLDGDWLLKFGYPLRSLLSPVFRSMQGVRRFISDPK 186
 b 130 RFLAYSEAVFAE---PEEGYLKLGAVFPFLEFSSMKAAAPKLVRLEAMRSVYVMSRERDHH 186
 Y 187 LVEILNYFIKVGSSPYDAPALMNLPIYIOHYGLAVVKGMYGMAQAMEKLAVELGVEI 246
 b 187 LRQAFSFLVLVGNPFATSIYTLIHALERNGWVFPFGGTGALVQGMVKLFEDLGG 246
 Y 247 RLDAEVESEIQKQGRACAVKLANGDVLADIVVSNMEVIPAMEKILRSPA---SELKMQ 303
 b 247 RLTSFVDEIVWEGQRTAVKLSGHTLPFDIVASNADVVHTYRELLRGAARSGRSGARLA 306
 Y 304 RFEPSCGLVHLGVDRLYPOLAHNFFYSDHPRHFDAVFKSHRLSDDPPIYLVAPCKT 363
 b 307 KTRHSMNLSFVIYFELNHHNQLAHHTVCFPRYRELIDEIFNKDGLAEDFSLYLHAPC 366

```
364 DPAQAGCEIILKILPHILPHLDKLLTADYSAERVLVVKLRMGLTDLRQHVITSEY 423
367 DKSLAPEGCTAFVVLSPFVHLGKADIDWEVGGPLVYDRDRIILAHLEERLLPGLDLSVTSRI 426
424 WPLDIOAKYNSNGSIYGVVADRFXNLGFKAPQSSSELSNLYFYVGGSNVPGGEMVTL 483
427 LTFPGRFDELSAFCGSAFSEPIILTSQAFWRPHNRDKANLYFAGATHPGAGVGGVVG 486
484 SGLVRDKIVADL 496
487 SAKATAGLILADL 499

RESULT 10
8GCSI PRELIMINARY; PRT; 492 AA.
C QBGCSI;
T 01-MAR-2003 (T-EMBLrel. 23, Created)
T 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Phytoene desaturase.
N CRTI.
S Pantoea stewartii.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Pantoea.
X NCBI_TaxID=66269;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 8200;
T deSouza M.L., Kollmann S.R., Schroeder W.A.;
L "Carotenoid Biosynthesis (WO 02/079395 A2).";
R Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
R EMBL: AY166713; AA85599.1; -.
R GO: GO:0016491; F:oxidoreductase activity; IEA.
R GO: GO:0009058; P:biosynthesis; IEA.
R GO: GO:0006118; P:electron transport; IEA.
R InterPro: IPR002937; Amino oxidase.
R Pfam: PF01593; Amino oxidase.
R PROSITE: PS00982; PHYTOENE DH; 1.
Q SEQUENCE 492 AA; 54836 MW; F3DCD224547A5FBA CRC64;

Query Match 26.4%; Score 687; DB 2; Length 492;
Best Local Similarity 33.3%; Pred. No. 1.2e-45;
Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

364 DPAQAGCEIILKILPHILPHLDKLLTADYSAERVLVVKLRMGLTDLRQHVITSEY 423
367 DKSLAPEGCTAFVVLSPFVHLGKADIDWEVGGPLVYDRDRIILAHLEERLLPGLDLSVTSRI 426
424 WPLDIOAKYNSNGSIYGVVADRFXNLGFKAPQSSSELSNLYFYVGGSNVPGGEMVTL 483
427 LTFPGRFDELSAFCGSAFSEPIILTSQAFWRPHNRDKANLYFAGATHPGAGVGGVVG 486
484 SGLVRDKIVADL 496
487 SAKATAGLILADL 499

RESULT 11
Q47845 PRELIMINARY; PRT; 492 AA.
AC Q47845;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTI.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
N [1]
P SEQUENCE FROM N.A.
R MEDLINE=93371414; PubMed=8395826;
R Liu S.T.;
R "Carotenoid-biosynthesis genes as a genetic marker for the purpose of
R gene cloning.";
R Biochem. Biophys. Res. Commun. 195:259-263(1993).
N [2]
P SEQUENCE FROM N.A.
R MEDLINE=94236237; PubMed=8180698;
R To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
R Chang Y.S., Liu S.T.;
R "Analysis of the gene cluster encoding carotenoid biosynthesis in
R Erwinia herbicola Ehol3.";
R Microbiology 140:331-339(1994).
R EMBL: M90698; AAA21263.1; -.
R PIR: S52586; S52586.
R GO: GO:0016491; F:oxidoreductase activity; IEA.
R GO: GO:0009058; P:biosynthesis; IEA.
R GO: GO:0006118; P:electron transport; IEA.
R InterPro: IPR002937; Amino oxidase.
R InterPro: IPR008150; Bac_phytoene_dh.
R InterPro: IPR008151; Phyt_n_dehydro.
R Pfam: PF01593; Amino oxidase; 1.
R ProDom: PD139017; Phyt_n_dehydro; 1.
R PROSITE: PS00982; PHYTOENE DH; 1.
Q SEQUENCE 492 AA; 55010 MW; 2D65E1A2A32D0635 CRC64;

Query Match 26.4%; Score 685; DB 2; Length 492;
Best Local Similarity 33.3%; Pred. No. 1.8e-45;
Matches 164; Conservative 90; Mismatches 226; Indels 12; Gaps 5;

11 VIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLMITKDGTFDLGPSILTMPIHFEA 70
6 VIGAGFGLALALRLQAAGIPVLLSQRKPGGRAVYEDQGFEDAGPTVITDPSAIE 65
71 LFTGAGKNMADYVQIQKVEPHWRNFFDGSVIDLCDAETQRRELKLGPGTYAQFREL 130
66 LFTLAGKQLADYVELLPVTFYRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVEGRQFL 125
131 DYSKNLCTETEGYFAKGLDGFWDLLKFGYPLRS---LLSPDVFERSMDQGVRRFISDPKL 187
126 DYSAVFE---GYLKLGTVPF---LSFRDLRAAPQLAKLQAWRTVYSKVASYIEDEHL 179
188 VEILNYIKTVYVGGSPYDAPALMNLPIYQHYGLWTVYKGMGMQAAMEKLAIVELGVEIR 247
180 RQAFSPHLLVGGNPFATSIYTLIHALEREWGVPFRGCTGALVNGMIKLFQDLGGEVV 239
248 LDAEVSIOQDGRACAVKLANGDVLPAIVSNMEVIPAMEKLL-RSPAS--ELKQKQR 304
```

```
360 PSLAPEGCSYVYLAPVPHLGTANLDWAVEGPRLRDRIFDYLEQHYMPGLRSQLVTHRMF 419
425 TPLDIOAKYNSNGSIYGVVADRFXNLGFKAPQSSSELSNLYFYVGGSNVPGGEMVTL 484
420 TPFDFRDELNAMQGSFAFSEPIILTSQAFWRPHNRDKANLYLVGAGTHPGAGIPGVIGS 479
485 GOLVRDKIVADL 496
480 AKATAGLILADL 491

RESULT 11
Q47845 PRELIMINARY; PRT; 492 AA.
AC Q47845;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTI.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
N [1]
P SEQUENCE FROM N.A.
R MEDLINE=93371414; PubMed=8395826;
R Liu S.T.;
R "Carotenoid-biosynthesis genes as a genetic marker for the purpose of
R gene cloning.";
R Biochem. Biophys. Res. Commun. 195:259-263(1993).
N [2]
P SEQUENCE FROM N.A.
R MEDLINE=94236237; PubMed=8180698;
R To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,
R Chang Y.S., Liu S.T.;
R "Analysis of the gene cluster encoding carotenoid biosynthesis in
R Erwinia herbicola Ehol3.";
R Microbiology 140:331-339(1994).
R EMBL: M90698; AAA21263.1; -.
R PIR: S52586; S52586.
R GO: GO:0016491; F:oxidoreductase activity; IEA.
R GO: GO:0009058; P:biosynthesis; IEA.
R GO: GO:0006118; P:electron transport; IEA.
R InterPro: IPR002937; Amino oxidase.
R InterPro: IPR008150; Bac_phytoene_dh.
R InterPro: IPR008151; Phyt_n_dehydro.
R Pfam: PF01593; Amino oxidase; 1.
R ProDom: PD139017; Phyt_n_dehydro; 1.
R PROSITE: PS00982; PHYTOENE DH; 1.
Q SEQUENCE 492 AA; 55010 MW; 2D65E1A2A32D0635 CRC64;

Query Match 26.4%; Score 685; DB 2; Length 492;
Best Local Similarity 33.3%; Pred. No. 1.8e-45;
Matches 164; Conservative 90; Mismatches 226; Indels 12; Gaps 5;

11 VIGAGLGLSAAISLATAGFSVOLIEKNDKVGKLMITKDGTFDLGPSILTMPIHFEA 70
6 VIGAGFGLALALRLQAAGIPVLLSQRKPGGRAVYEDQGFEDAGPTVITDPSAIE 65
71 LFTGAGKNMADYVQIQKVEPHWRNFFDGSVIDLCDAETQRRELKLGPGTYAQFREL 130
66 LFTLAGKQLADYVELLPVTFYRLCWESGKVFYNDNDQAQLEAQIQQFNPRDVEGRQFL 125
131 DYSKNLCTETEGYFAKGLDGFWDLLKFGYPLRS---LLSPDVFERSMDQGVRRFISDPKL 187
126 DYSAVFE---GYLKLGTVPF---LSFRDLRAAPQLAKLQAWRTVYSKVASYIEDEHL 179
188 VEILNYIKTVYVGGSPYDAPALMNLPIYQHYGLWTVYKGMGMQAAMEKLAIVELGVEIR 247
180 RQAFSPHLLVGGNPFATSIYTLIHALEREWGVPFRGCTGALVNGMIKLFQDLGGEVV 239
248 LDAEVSIOQDGRACAVKLANGDVLPAIVSNMEVIPAMEKLL-RSPAS--ELKQKQR 304
```


128 RFLD--YSKNLCTETETAGYF-----AKGLDGFWDLLKFKYGLRSLLSFDFVFRSM---DQG 177
122 SFLTDVYK--YEIARRYFLERTYKPSD-----FYN-MTSLVQGAKLKTLNHAQDL 170
178 VRFPIEDPKLVEITLNYFIKVGSSPYDAPALMNLPLVIOYHYGLMYKGMVGMQAQMEK 237
171 IEHYIDNEKIQLKLAFTQLYIGIDPKRGPSLSYIIPMIEMFGVHFIKGMVGMQAQGLAQ 230
238 LAVELGVEIRLDAEVS--IQKQDGRACAVKLANGDVLDPADIVVSNMEVIPAMEKLRSP 295
231 LNKDLGWNIELNAIEQIILIDPKFKADAIKV-NGDIRKFDKILCTADPPSVAESLMPDF 289
296 ASEK-----KQORFEPSCSLVHLGVD-RLYPOLAHNFFYSDHREHFDVAFKSHRLS 350
290 APIKKYPPHKAIDLVSCLSAFLMYIGIDVTDQVRLHNVIFSDDPGRNIEIFEG-RLS 348
351 DDPTIYLVAPCKTDPQAQAGCBIILKILPHILD-----PKLLTAEDYSALRERV 403
349 YDPSIYVYVPAVADKSLAPGKGIYVLMPTPELKTGSGIDMSDEALTQO-----IKEIIV 404
404 VKLERMGL-TDLRQHIVTEYWTPLDIQAKYYSNQSIIYGVVADRFRKFLGFKAPQSSSEL 462
405 RKLATIEVFEDIKSHIVSETIFTENDFEQTYHAKFGSAFGLMPTLAQSNYYRPNQVSRDY 464
463 SNLYFVCGSVNPGGMPMTLSGLVDRDKIVADLQ 497
465 KDLFYAGASTHFGAGVPIVLTSAKIIVDEMICKDIE 499

RESULT 14
007855 PRELIMINARY; PRT; 506 AA.
NC 007855; 053723;
DT 01-JUN-1997 (TrEMBLrel. 04, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Squalene synthase.
GN CRTN.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Newman;
RA MEDLINE=95095943; PubMed=8002598;
RA Wieland B., Feil C., Gloria-Waerker E., Thumm G., Lechner M.,
RA Bravo J.M., Poralla K., Goetz F.
RT "Genetic and biochemical analyses of the biosynthesis of the yellow
carotenoid 4,4'-diaponeurosporene of Staphylococcus aureus.";
RL J. Bacteriol. 176:7719-7726 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Newman;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBSJ databases.
RL EMBL; X73889; CAA52098.1; -
DR GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000759; Adrnx reductase.
DR InterPro; IPR002937; Amino Oxidase.
DR InterPro; IPR000205; NAD BS.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXEDTASE.
DR ProDom; PD139017; Phytin dehydro; 1.
SQ SEQUENCE 506 AA; 57231 MW; 3CA38CB6C5668B49 CRC64;
Query Match 25.7%; Score 667; DB 2; Length 506;
Best Local Similarity 31.3%; Pred. No. 4.9e-44;
Matches 161; Conservative 120; Mismatches 192; Indels 42; Gaps 14;
8 RVIVTGAGLGGLSAISAISATAGFSVOLLKXNDKVGKGLNITWKGFTFDLGPSTLTPHI 67

Db 2 KIAIAGVGTGLAAARIAIOGHEVTIPEKNNVGRMNQLKKGDFTFDMGPTIIVMPDV 61
Qy 68 FEALFTGAGKNWADYVQIKVPHWFNFEDSGVLDLCEDAETQRELDKLGPGYVAQFO 127
Db 62 YKDFVFTACGVKNVEDYIELRQLRYIVDYVFDHDDRITVPTDLAELQOMLESIEFGSTHGM 121
Qy 128 RFLD--YSKNLCTETETAGYF-----AKGLDGFWDLLKFKYGLRSLLSFDFVFRSM---DQG 177
Db 122 SFLTDVYK--YEIARRYFLERTYKPSD-----FYN-MTSLVQGAKLKTLNHAQDL 170
Qy 178 VRFPIEDPKLVEITLNYFIKVGSSPYDAPALMNLPLVIOYHYGLMYKGMVGMQAQMEK 237
Db 171 IEHYIDNEKIQLKLAFTQLYIGIDPKRGPSLSYIIPMIEMFGVHFIKGMVGMQAQGLAQ 230
Qy 238 LAVELGVEIRLDAEVS--IQKQDGRACAVKLANGDVLDPADIVVSNMEVIPAMEKLRSP 295
Db 231 LNKDLGWNIELNAIEQIILIDPKFKADAIKV-NGDIRKFDKILCTADPPSVAESLMPDF 289
Qy 296 ASEK-----KQORFEPSCSLVHLGVD-RLYPOLAHNFFYSDHREHFDVAFKSHRLS 350
Db 290 APIKKYPPHKAIDLVSCLSAFLMYIGIDVTDQVRLHNVIFSDDPGRNIEIFEG-RLS 348
Qy 351 DDPTIYLVAPCKTDPQAQAGCBIILKILPHILD-----PKLLTAEDYSALRERV 403
Db 349 YDPSIYVYVPAVADKSLAPGKGIYVLMPTPELKTGSGIDMSDEALTQO-----IKEIIV 404
Qy 404 VKLERMGL-TDLRQHIVTEYWTPLDIQAKYYSNQSIIYGVVADRFRKFLGFKAPQSSSEL 462
Db 405 RKLATIEVFEDIKSHIVSETIFTENDFEQTYHAKFGSAFGLMPTLAQSNYYRPNQVSRDY 464
Qy 463 SNLYFVCGSVNPGGMPMTLSGLVDRDKIVADLQ 497
Db 465 KDLFYAGASTHFGAGVPIVLTSAKIIVDEMICKDIE 499

RESULT 15
QSHPD8
ID Q9HPD8 PRELIMINARY; PRT; 512 AA.
AC Q9HPD8;
DT 01-MAR-2001 (TrEMBLrel. 15, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTII OR VNGI684G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AE005076; AAG19932.1; -
DR PIR; H84320; H84320.
DR GO:0006118; P:electron transport; IEA.
DR GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000759; Adrnx reductase.
DR InterPro; IPR002937; Amino Oxidase.
DR InterPro; IPR001327; FAD_Pyr_redox.
DR InterPro; IPR000205; NAD BS.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXEDTASE.

```
{} PRINTS; PRO0368; FADPNR.
{} ProDom; PDI39017; Phytin dehydro; 1.
{} FAD; Flavoprotein; Oxidoreductase; Complete proteome.
{} SEQUENCE 512 AA; 56274 MW; 95C75E5A05F9126 CRC64;

Query Match      25.5%; Score 662.5; DB 17; Length 512;
Best Local Similarity 32.4%; Pred. No. 1.1e-43;
Matches 169; Conservative 81; Mismatches 217; Indels 55; Gaps 11;

9 VIVIGAGLGGLSAISLATAGFSVOLIEKNKVGKLNIMTKDGTFTDLGPSILTMPIHF 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9 VAVIGAGFGGLSTACYLADAGADVTVVEKTDQIGGRASTLERDGRFDGMPGYLMPDVF 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

69 EALFTGAGKNMADYVQICKVSPHNEFE-----DGSVIDLCEDAET 110
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 EEPFASFSHPFSDYGLTRDPHRIFFKONEGRPGEDAPGLNVDTDGTIDVTPDREQ 128
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

111 QRRELDKLGPGTYAQFQFLPYSKNECTETBAGYFAKGLDGF-----WDLKPYG 160
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
129 VKQVFDAVEPGAGDVLDDYLAQAKE-----NVEVGMEHFVKTDRPRVDRWMDPKLAE 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

161 PLSRLSFDVFRSMDOGYRRFISDPKVEILNYPKYVGSSPYDAPALMNLPLVIOHYG 220
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 YARGL---TLGSMQDHVEQYFDRPKQQVQYTLVFLGGSPTTTPALYNLMSHVDFGLG 237
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

221 LMVYKGMVGMQAAMEKLAIVELGVEIRLDAEVSEIQODGRACAVKLAN--GDVLPADIV 278
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 VYYPGGMGVVVDGITALARELGVEFRGTGHPVSAIK--GRRGGFKLDTBPGDAVLADV 294
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

279 VSNMEVITAMEKLL-----RSPASELKMQRFEPSCSGLVHLGVDRLYPOLAHENFFYSD 334
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 VSDADYAHTEQALLPAKQKQVDADYWDARTYAP--SAFLLYLGVGDEVEPIAHTLVLP 352
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

335 HPREHFDVAFKSHRLSDPTIYLVAPCKTDPAQAPAG-CEIILPHIPHLDPDKLITAE 393
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 DWDTHFAQIFDDPAPNPDPAAYLVCVPSKTDVTAPDGHSLFALVPVAPGLD-DTPAVRE 411
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

394 DYSALRERVVLKRLKMLTDLRQHLVTEEYWTPLDIOAKYYSNQGSIYGVVADRFKNLGF 453
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 SY---RDLVLDDIAENTGVLDLDRIVVEERFSVSEFADRYNSHQGTALGLAHTLROTALF 468
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

454 KAPQSSSELNLYFVGGSVNPGGMPVTLGQLVRDKIVAD 495
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 RPPHRSALDGLYFTGANTTPGIGVPMCLISGEVTANNVLAD 510
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: February 25, 2004, 14:50:45
Job time: 42.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 8.3764 Seconds
(without alignments)
3089.496 Million cell updates/sec

File: US-09-941-947A-24

Effect score: 2598

Sequence: 1 MNSNDNRVIVIGAGLGSL.....MPWVTLGGQLVRDKIVADLIQ 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARY25

Result No.	Score	Query Match	Length	ID	Description
1	773	29.8	517	1 CRTJ MYXXA	P54979 Myxococcus
2	770	29.6	492	1 CRTJ ERWHE	P22871 erwinia her
3	671	25.8	492	1 CRTJ PANAN	P21685 pantoaea ana
4	645	24.8	501	1 CRTJ AGRAU	P54978 agrobacteri
5	641.5	24.7	583	1 CRTJ PHYBL	P54982 phytoce
6	638	24.6	529	1 CRTJ MYXXA	Q02861 myxococcus
7	631.5	24.3	621	1 CRTJ CERNX	P48537 cercospora
8	624.5	24.0	595	1 CRTJ NEUCR	P21334 neurospora
9	581.5	22.4	524	1 CRTJ RHOCB	P17054 rhodobacter
10	548	21.1	518	1 CRTJ RHOSH	P54980 rhodobacter
11	541.5	20.8	495	1 CRTD RHOSH	Q01671 rhodobacter
12	526.5	20.3	507	1 CRTJ STRGR	P54981 streptomyce
13	514.5	19.8	508	1 CRTJ STRSE	P54971 streptomyce
14	490.5	18.9	494	1 CRTD RHOCB	P17059 rhodobacter
15	174.5	6.7	472	1 CRTJ SYN3	P29273 synchocyst
16	158.5	6.1	570	1 CRTJ NARPS	Q40406 narcessus p
17	157	6.0	516	1 OXLA CROAD	Q93364 crotalus ad
18	156	6.0	519	1 AOFB RAT	P13643 rattus norv
19	152.5	5.9	570	1 CRTJ SOYEN	P28553 glycine max
20	150	5.8	500	1 PAO MAIZE	Q64411 zea mays (m
21	149.5	5.8	566	1 CRTJ ARATH	Q07356 arabidopsis
22	148.5	5.7	474	1 CRTJ SYN7	P26294 synchococc
23	146	5.6	461	1 P49 STPLI	Q06108 streptomyce
24	145.5	5.6	535	1 Y897 MYCTU	Q10555 mycobacteri
25	144.5	5.6	478	1 PUO MICRU	P40374 micrococcus
26	144	5.5	500	1 PCLL ARATH	P57681 arabidopsis
27	143.5	5.5	571	1 CRTJ MAIZE	P49086 zea mays (m
28	142.5	5.5	566	1 CRTJ ORYSA	Q92tn9 oryza sativ
29	142	5.5	508	1 FMS1 YEAST	P50264 saccharomyc
30	139.5	5.4	583	1 CRTJ LYCES	P28554 lycopersico
31	136.5	5.3	522	1 AOF ONCMY	P49253 oncorhynch
32	135	5.2	544	1 A37C DROLE	Q96570 drosophila
33	134	5.2	630	1 FIGI MOUSE	Q09046 mus musculu

ALIGNMENTS

RESULT 1

ID	CRTJ MYXXA	STANDARD;	PRT;	517 AA.
AC	P54979;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).			
GN	CARA2.			
OS	Myxococcus xanthus.			
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;			
OX	Cystobacterineae; Myxococcaceae; Myxococcus.			
NCBI_TaxID=34;				
[1]				
SEQUENCE FROM N.A.				
STRAIN=DK1050;				
MEDLINE=96061955; PubMed=7588751;				
Botella J.A., Murillo F.J., Ruiz-Vazquez R.M.;				
"A cluster of structural and regulatory genes for light-induced				
carotenogenesis in Myxococcus xanthus.";				
Eur. J. Biochem. 233:238-248(1995).				
CC	FUNCTION: This enzyme converts phytoene into zeta-carotene via the			
CC	intermediary of phytofluene by the symmetrical introduction of two			
CC	double bonds at the C-11 and C-11' positions of phytoene.			
CC	COFACTOR: FAD (Probable).			
CC	PATHWAY: Carotenoid biosynthesis.			
CC	SIMILARITY: Belongs to the phytoene dehydrogenase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; Z21955; CAA79956.1; -			
DR	PIR; S32169; S32169.			
DR	InterPro; IPR000759; Adrndx_reductase.			
DR	InterPro; IPR002937; Amino oxidase.			
DR	InterPro; IPR008150; Bac.phytoene_dh.			
DR	InterPro; IPR000205; NAD_BS			
DR	InterPro; IPR008151; Phyt_n_dehydro.			
DR	Pham; PFO1593; Amino oxidase; 1.			
DR	PRINTS; PR00419; ADXRUTASE.			
DR	ProDom; PD139017; Phyt_n_dehydro; 1.			
DR	PROSITE; PS00982; PHYTOENE_DH; 1.			
DR	Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.			
FT	NP BIND. 11 44 FAD (ADP PART) (POTENTIAL).			
SEQUENCE	517 AA; 57783 MW; CECDD74AB9F9P8CB CRC64;			

Query Match 29.8%; Score 773; DB 1; Length 517;

Best Local Similarity 34.7%; Pred. No. 2.2e-51;

Matches 174; Conservative 99; Mismatches 205; Indels 24; Gaps 9;

QY 3 SNDNRVIVIGAGLGSLAAISLATAGFSVQLIKNDKVGKLNIMTKDGTDFDLGPSIL 62

DR EMEL; M87280; AAA64981.1; --
DR PIR; A39273; A33120.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phyt. dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR Prodom; PD139017; Phyt. dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; PAD; Flavoprotein; NAD.
FT NP BIND 5 38 PAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8BDC5DB1562083F2 CRC64;

Query Match 29.6%; Score 770; DB 1; Length 492;
Best Local Similarity 34.7%; Pred. No. 3.5e-51;
Matches 174; Conservative 93; Mismatches 213; Indels 22; Gaps 6;

QY 7 QRVIVGAGLGGSLAALSLATAGSVOLIEKNKVGKLNIMTGGTDLGSLTMDPH 66
DB 2 KKTWVIGAGFGGLALAIRLQAAGIPTVLLQEQKGGRAVYVWHDQGGFTDAGPTVITDPT 61

QY 67 IFEALFTGAGKGMADYVQIQVBPHEWNEFFEDGSDVIDLCEDAETORRELDKLGFTYVAF 126
DB 62 ALEALFTLAGRMEDYVRLLPVKFYRLCWESGKTLIDYANDSAELEAQITQFNPRDVEGY 121

QY 127 QRFIDYSKNLCTETAGYFAXG---LDGFWDLKFKYQPLRSLLSFDVFRSMDOGVRRFIS 183
DB 122 RFLAYSQAVFOE---GYLRGVSVPFLSFRDMLR-AGP---QLLKLAQWSYQSVSRFIE 175

QY 184 BPKLVILNLYIKVVGSPYDAPALMNLPLYOYHGLVYVGMGMGAQAMEKLAVELG 243
DB 176 DEHLRQAFPSHLLVGGNPFPTSSITLILHALEREMGVWPEGGTGALVNGVXKLPDGLG 235

QY 244 VEIRLDAEVSIEIQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSP 295
DB 236 GSEILNARVEELVADNRVSOVRLADGRI FDTDAVASNADVNTYKLLGHHHPYQKRAA 295

QY 296 ASELKQORFEPSCSLVHLGVDRLYPOLAHNFFYSDHPRHFDAVERSHRSLDDPTI 355
DB 296 ALERKSM-----SNSLFVLYFGLNPHSQLAHHTICFPYRELIDEIFTGSAADDFSL 350

QY 356 YLVAPCKTDPAPAGACCEIILPHIPHLDPDKLLTAEDYSALRERVVLKLERMGLTDLR 415
DB 351 YLHSTCVTDPSLAPPGCAFVFLAPVPHLGNAPLDWAQEGKPLDRIFDYLEERYMPGLR 410

QY 416 QHIVTEEYWTPLDIOAKYNSQSGIYGVVADRFKMLGFKAPORSSELSNLYFYGVSNPG 475
DB 411 SOLVTRQITFPADFHDTLDALHGSFAFSIEPLLTQSAMFRPHNRDSIDIANLYLVGATHPG 470

QY 476 GGPWVTLGGQLVRDKIVADIQ 497
DB 471 AGIPGVVASAKATASLMIEDLQ 492

RESULT 3
CRTL PANAN
ID CRTL PANAN STANDARD; PRT; 492 AA.
AC P22871;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTL
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20D3;
RX MEDLINE=91072214; PubMed=2254247;
RA Misawa N.; Nakagawa M.; Kobayashi K.; Yamano S.; Izawa Y.;
RA Nakamura K.; Harashima K.;
RT "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway

EMEL; M38423; AAA24820.1; --
DR PIR; A39273; A33120.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phyt. dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR Prodom; PD139017; Phyt. dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; PAD; Flavoprotein; NAD.
FT NP BIND 5 38 PAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8BDC5DB1562083F2 CRC64;

Query Match 29.6%; Score 770; DB 1; Length 492;
Best Local Similarity 34.7%; Pred. No. 3.5e-51;
Matches 174; Conservative 93; Mismatches 213; Indels 22; Gaps 6;

QY 7 QRVIVGAGLGGSLAALSLATAGSVOLIEKNKVGKLNIMTGGTDLGSLTMDPH 66
DB 2 KKTWVIGAGFGGLALAIRLQAAGIPTVLLQEQKGGRAVYVWHDQGGFTDAGPTVITDPT 61

QY 67 IFEALFTGAGKGMADYVQIQVBPHEWNEFFEDGSDVIDLCEDAETORRELDKLGFTYVAF 126
DB 62 ALEALFTLAGRMEDYVRLLPVKFYRLCWESGKTLIDYANDSAELEAQITQFNPRDVEGY 121

QY 127 QRFIDYSKNLCTETAGYFAXG---LDGFWDLKFKYQPLRSLLSFDVFRSMDOGVRRFIS 183
DB 122 RFLAYSQAVFOE---GYLRGVSVPFLSFRDMLR-AGP---QLLKLAQWSYQSVSRFIE 175

QY 184 BPKLVILNLYIKVVGSPYDAPALMNLPLYOYHGLVYVGMGMGAQAMEKLAVELG 243
DB 176 DEHLRQAFPSHLLVGGNPFPTSSITLILHALEREMGVWPEGGTGALVNGVXKLPDGLG 235

QY 244 VEIRLDAEVSIEIQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSP 295
DB 236 GSEILNARVEELVADNRVSOVRLADGRI FDTDAVASNADVNTYKLLGHHHPYQKRAA 295

QY 296 ASELKQORFEPSCSLVHLGVDRLYPOLAHNFFYSDHPRHFDAVERSHRSLDDPTI 355
DB 296 ALERKSM-----SNSLFVLYFGLNPHSQLAHHTICFPYRELIDEIFTGSAADDFSL 350

QY 356 YLVAPCKTDPAPAGACCEIILPHIPHLDPDKLLTAEDYSALRERVVLKLERMGLTDLR 415
DB 351 YLHSTCVTDPSLAPPGCAFVFLAPVPHLGNAPLDWAQEGKPLDRIFDYLEERYMPGLR 410

QY 416 QHIVTEEYWTPLDIOAKYNSQSGIYGVVADRFKMLGFKAPORSSELSNLYFYGVSNPG 475
DB 411 SOLVTRQITFPADFHDTLDALHGSFAFSIEPLLTQSAMFRPHNRDSIDIANLYLVGATHPG 470

QY 476 GGPWVTLGGQLVRDKIVADIQ 497
DB 471 AGIPGVVASAKATASLMIEDLQ 492


```

b 309 WSNLSVLHGLREAPKDVAAHTILFGPRYKELVNEIFKPGKLAEDPSLYLHSPCTTDE 368
y 367 QAPAGCEIIKILPHILDPKLLTAEDYSALRERVLVGLERMLTDLRQHVITEYWT 426
b 369 MAPPGKSTHYVLAPVPHGLRADIDWAVSGPRYADRIILASLEERLIPLNLANLTTRIFTP 428
y 427 LDIQAKYYSNOSIYGVVADRFKNLGPAPORSSSELNLYFVGSVNPGGMPWVTLSCQ 486
c 429 SDFASELANHGSASFSEIITQSANFRPHNRDRTIRNFYLVGAGTHPGAGIGPVVGS 488
y 487 LVRDKIVADL 496
b 489 ATAQVWLSDL 498

RESULT 5
RTI_PHYBL
D CRTI_PHYBL STANDARD; PRT; 583 AA.
C P54982;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CARB.
S Phycomyces blakesleeanus.
C Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
C Phycomyces.
X NCBI_TaxID=4837;
P SEQUENCE FROM N.A.
C STRAIN-NRRL 1555;
A Ruiz-Hidalgo, M.J.;
L Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
C -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
C -!- COFACTOR: FAD (Probable).
C -!- PATHWAY: Carotenoid biosynthesis.
C -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
C
C EMBL; X78434; CAA55197.1; -
C PIR; S43139; S43139.
C InterPro; IPR002937; Amino oxidase.
C InterPro; IPR008150; Bac_phytoene_dh.
C InterPro; IPR00205; NAD_B5.
C Pfam; PF01593; Phytn dehydro.
C ProDom; PD139017; Phytn dehydro; 1.
C PROSITE; PS00982; PHYTOENE_DH; 1.
C Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
C Transmembrane.
C NP_BIND 8 41 FAD (ADP PART) (POTENTIAL).
C TRANSMEM 531 551 POTENTIAL.
C SEQUENCE 583 AA; 65983 MW; 60E8F682B12FB591 CRC64;
Query Match 24.7%; Score 641.5; DB 1; Length 583;
Best Local Similarity 31.3%; Pred. No. 2.8e-41;
Matches 157; Conservative 100; Mismatches 223; Indels 21; Gaps 11;
7 QRVIVIGAGLGLSAAISLAFAGSVGLISKNOKVGGKLNIMTKDGTFTDLGFSILTWPH 66
5 KHVIIIGAGGTATAFLAREGKIVTVSKNFFGGRCSLINHGHRFDQGPSLYLMPK 64
67 IFEALFTGAGKNADYYQIOKVEPHWENFFEDGSDVLDLCEADTAQRELDKL-GPGTVAQ 125

```

```

Db 65 LFEAFEPALDEKIEDHVELLRCHNNYKVHFDGDKIQSSLSRMKPMERIEGPD---G 121
Cy 126 FQRFELDYSKMLCTETETAGY---PAKGLDGFMDLLKF-YCPILRSLSLSPVFRSMQGVRRF 181
Db 122 FLRFELDMKESHYEGGVEMAIKONFETIKWLRLQVVP-ALFELHIFDFVYSRAAKY 179
Cy 182 ISPKLVELILNYPIKVVSSPYDAPALANLLIPIYQVHYGLA-VKGMKYGMAQMEKLAIVE 241
Db 180 FTKKMMMAFTPOSTWNGSPYDSPAIVNLLQYTFASGIR-PRGQFNVTYOKLENIAIE 239
Cy 242 -LGVEIRLDAEYSEIQKD--GRACAVKLANGDVLPAIVV-NNMEVIFAMEKLLNSPASE 298
Db 240 KFGARFIYEAPVAKINTDDKGGKVTGVTIQSGEVEADAVV-NADLVYAVHNLPPCSEWT 299
Cy 299 LKKMQRFEPSCGLVHLGVDRILYPOLAHENFFYSDHREHFDVAFKSHRLSDDDTIYLV 358
Db 300 TNLIAEKKLHSSSISFYMSLKKVWPELDVHNIFLAFAFKES-DEIFTDKWPELSFYVN 359
Cy 359 APCKTDPAPAGACEIIKILPHIPLHDPKLLTADYSAL-----RERVVLKLE-RMGUTD 413
Db 360 LPSRIDPTAAPPGKOSMIVLVPIGHM-KSKTWEADYTWIVKARKMVLVLELRRLGLTN 418
Cy 414 LRQHIVTEEVWTPLDIOAKYYSNOSIYGVVADRFKNLGFKAPORSS--ELSNLYFVGS 471
Db 419 FIDLVHEEHWNDPSIQKKFNLWRGSLILGLSHDVLQVLFNRPSTQDSTGRYKNLFFVGAS 478
Cy 472 VNPGGMPWVTLSCQLVRDKI 492
Db 479 THPGTGVPVLAGSKLTSQV 499

RESULT 6
CRTI_MXXA
ID CRTI_MXXA STANDARD; PRT; 529 AA.
AC Q02861;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
N CARC.
S Myxococcus xanthus.
C Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
C Cystobacterineae; Myxococcaceae; Myxococcus.
X NCBI_TaxID=34;
P SEQUENCE FROM N.A.
C STRAIN=NR403;
C MEDLINE=93223667; PubMed=8467787;
C Fontes M., Ruiz-Vazquez R.M., Murillo F.J.;
C "Growth phase dependence of the activation of a bacterial gene for
C carotenoid synthesis by blue light.";
C EMBO J. 12:1265-1275(1993).
C -!- FUNCTION: This enzyme converts phytoene into lycopene via the
intermediaries of phytofluene, zeta-carotene and neurosporene by
C the introduction of four double bonds (by similarity).
C -!- COFACTOR: FAD (Probable).
C -!- PATHWAY: Carotenoid biosynthesis.
C -!- INDUCTION: By blue light.
C -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
C
C EMBL; M94727; AAA25390.1; -
C PIR; S35306; S35306.
C InterPro; IPR000759; Adrndx reductase.
C InterPro; IPR002937; Amino Oxidase.
C InterPro; IPR008150; Bac_phytoene_dh.

```

RESULT 8

5

1.A.

199

```
1; PubMed=3085820;
```

Daub M.E.;

quene, and chalcone dehydrogenase

Microbiol. 60:2766

Db	228	PPGGPHKVLDAVKIGERMVGKYLRLNTGVSVQLTDCGKNGKPKATGVQLNGEVLNADL	287
Qy	278	VYSNNEVIPAMEKILRLSPASELK-----KMQRFSPSCGLVHLGVDRLVPQLAHENFFVS	333
Db	288	VVNADLVVYTNVLLPKIEIGIKKYANKLANRKASCSSISFYWLSLGMKAKELETENIFLA	347
Qy	334	DHPREHEDAVKSHRLSDDDTYLVAPCKTDPAQAPAGCE-IKILP--HI-----PHLD	385
Db	348	REYKSPDAIFERQALPDDPSFVTHVPSRVDPSPAADPRDAVIALVFGVGLLQNGQPELD	407
Qy	386	PKDLTADYDSALRERYLVKLE-RMGITDLRQHIVTEEYW-TPLDIAQKYYSNQSIIYGV	443
Db	408	WPTLV-----SKARAGVLATIQARTGLS--LSPLITEEIVNTFYTWETKFNLSKGAILGL	460
Qy	444	VADRFKNLGFKAQPSRSELSNEYFVGSGVNPGGMPVLTLSQLVDRDKIVAD	495
Db	461	AHDFENVLAFPRTYAQMGNDAIFVGASTHPGTGVPLVLAGAKITAEQILLEE	512
RESULT 9			
CRTI_RHOCA			
ID	CRTI_RHOCA	STANDARD;	PRT; 524 AA.
AC	P17054;		
DT	01-AUG-1990 (Rel. 15, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase).		
GN	CRTI.		
OC	Rhodobacter capsulatus (Rhodospseudomonas capsulata).		
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;		
OC	Rhodobacteraceae; Rhodobacter.		
OX	NCBI_TaxID=1061;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=SB1003 / St Louis, and BEC404;		
RC	MEDLINE=89313663; PubMed=2747617;		
RX	Armstrong G.A., Alberti M., Leach P., Hearst J.E.;		
RT	"Nucleotide sequence, organization, and nature of the protein		
RT	products of the carotenoid biosynthesis gene cluster of Rhodobacter		
RT	capsulatus";		
RL	Mol. Gen. Genet. 216:254-268 (1989).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=89327279; PubMed=2546948;		
RA	Bartley G.E., Scolnik P.A.;		
RA	"Carotenoid biosynthesis in photosynthetic bacteria. Genetic		
RT	characterization of the Rhodobacter capsulatus CrtI protein.";		
RL	J. Biol. Chem. 264:13109-13113 (1989).		
RP	[3]		
RP	ERRATUM.		
RA	Bartley G.E., Scolnik P.A.;		
RA	J. Biol. Chem. 264:18260-18260 (1989).		
RP	[4]		
RP	SIMILARITY TO CAROTENOID DESATURASES.		
RP	MEDLINE=90368827; PubMed=2144293;		
RA	Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;		
RT	"Carotenoid desaturases from Rhodobacter capsulatus and Neurospora		
RT	crassa are structurally and functionally conserved and contain		
RT	domains homologous to flavoprotein disulfide oxidoreductases.";		
RT	J. Biol. Chem. 265:16020-16024 (1990).		
CCC	-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the		
CCC	intermediary of phytofluene by the symmetrical introduction of two		
CCC	double bonds at the C-11 and C-11' positions of phytoene.		
CCC	-!- COFACTOR: FAD (probable).		
CCC	-!- PATHWAY: Carotenoid and chlorophyll biosynthesis.		
CCC	-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.		
CCC	----		
CCC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CCC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CCC	European Bioinformatics Institute. There are no restrictions on its		
CCC	use by non-profit institutions as long as its content is in no way		
CCC	modified and this statement is not removed. Usage by and for commercial		

entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
EMBL: J04969; AA50313.1; --
EMBL: X52291; CA336533.1; --
EMBL: Z11165; CA477540.1; --
PIR: A32617; A32617.
InterPro: IPR000759; Adrmdx_reductase.
InterPro: IPR002937; Amino oxidase.
InterPro: IPR008150; Bac_phytoene_dh.
InterPro: IPR000205; NAD_BS.
InterPro: IPR008151; Phyt_n_dehydro.
Pfam: PF01593; Amino oxidase; 1.
PRINTS: PR04419; ADXRTAS.
PRODOM: PD139017; Phyt_n_dehydro; 1.
PROSITE: PS00982; PHYTOENE_DH; 1.
Photosynthesis; Chlorophyll biosynthesis;
Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SEQUENCE 524 AA; 57978 MW; 5425A7B5A06A6B9 CRC64;

Query Match 22.4%; Score 581.5; DB 1; Length 524;
Best Local Similarity 29.5%; Pred. No. 9.2e-37;
Matches 151; Conservative 101; Mismatches 224; Indels 35; Gaps 10;

2 NSNDNORVIVIGAGLGLSAAISLATAGPSVOLIEKNDKVGKLNIMTKDGTFTDLQPSI 61
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 NTEGMRVAVIGAGLGLSAAISLATAGPSVOLIEKNDKVGKLNIMTKDGTFTDLQPSI 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

62 LTPGHLFEALFTGAGKQADYQIQVVEHWNFFEDGSVIDLCEADATQRELDKLGPG 121
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 VTVPDLRELWDGCRDFDROVSLVPMSEFFYITDFPDGERTATYDGDAAKVAEVARISPG 123
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

122 TYAQFORFLDYSKNLTCTEAGY---FAGLGDGFNDLKF---YGPLRSLLSFDVFRSMD 175
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 DVEGFRHFWDAK---ARVEFGYENLGRKPMKSLNDLIKVLPTFGWLR-----DRSVY 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

176 QGVRRISDEKLEILNYFIKYVGSPPDAPALMNLPIYIQTGLVTVKGMTCMAQAM 235
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 GHAKWVKDHLRFALSFLPFGIGGPHVTSMYILVSQLEKTKGVHYATGGVQAIADAM 234
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

236 EKLAVELGVEIRLDREVSI-QKQDGRACAVKLANGDVLPIVSNNEVIPAWEKLLRS 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
235 AKVITDQGGEMRLNTEVDEILVSRGKATGIRLMDGTLPAAQVVSNAADAGHTYKRLN 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

295 P-----ASELKQKQPEPSCSLVHL-----GVRLYQQLAHNFFVSDPREHDAVFK 345
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 RDRWRWTDKLDKKRMS---MGLFWYFETGKTAKQWQDVGHVTVVVGPRYKEHVQDIFI 351
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

346 SHRLSDPTIYLVAPCKTDPAPAGACEIKILPHIPLDPOK---LLTAEDYSALRR 401
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
352 KGLAEDMSLYVHRPSVTDPTAAPKGDFTFYVLPVNLGFDNGVDNSVEAEKYKA---X 408
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

402 VLKLERMGLTDLRQHIVTEETWFLDTQAKYNSQGSIVGVVADRPNKLGFKAPQSS 461
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
409 VLKVEERLLPGVAEKITEVVFTEFTRDRLYSLPLGAGFSLEPRILQSAWFRPHNASE 468
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

462 LSNLYFVGSGVNPFGGMPVTLSSGLQVRDKI 492
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
469 VDGLYLGVAGTHFGAGVSPVSGSELVAQMI 499
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
RTI RHOSH
C CRTI RHOSH STANDARD; PRT; 518 AA.
P P54980; Q9RFD0;
I 01-OCT-1996 (Rel. 34, Created)
I 01-OCT-1996 (Rel. 34, Last sequence update)
I 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
V CRTI.
S Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
S Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
```

```
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=94292403; PubMed=80211167;
RA Lang H.P., Cogdell R.J., Gardiner A.T., Hunter C.N.;
RT "Early steps in carotenoid biosynthesis: sequences and
RT transcriptional analysis of the crtI and crtB genes of Rhodobacter
RT sphaeroides and overexpression and reactivation of crtI in
RT Escherichia coli and R. sphaeroides.";
RL J. Bacteriol. 176:3859-3863(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=95238278; PubMed=7721699;
RA Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;
RT "Complete DNA sequence, specific Tn5 insertion map, and gene
RT assignment of the carotenoid biosynthesis pathway of Rhodobacter
RT sphaeroides.";
RL J. Bacteriol. 177:2064-2073(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=20115911; PubMed=10648776;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1."
RL Nucleic Acids Res. 28:862-867(2000).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
EMBL: S71770; AA831138.1; --
EMBL: AJ010302; CAB38739.1; --
EMBL: AF195122; AAF24289.1; --
PIR: S49620; S49620.
PIR: T50745; T50745.
DR InterPro: IPR002937; Amino oxidase.
DR InterPro: IPR008150; Bac_phytoene_dh.
DR InterPro: IPR000205; NAD_BS.
DR InterPro: IPR008151; Phyt_n_dehydro.
DR Pfam: PF01593; Amino oxidase; 1.
DR ProDom: PD139017; Phyt_n_dehydro; 1.
DR PROSITE: PS00982; PHYTOENE_DH; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 14 47 FAD (ADP PART) (POTENTIAL).
FT CONFLICT 73 73 R -> C (IN REF. 3).
FT CONFLICT 174 174 T -> S (IN REF. 3).
FT CONFLICT 292 292 L -> F (IN REF. 3).
FT CONFLICT 395 395 Q -> P (IN REF. 3).
SQ SEQUENCE 518 AA; 57244 MW; 423B3515639EF8F1 CRC64;

Query Match 21.1%; Score 548; DB 1; Length 518;
Best Local Similarity 28.3%; Pred. No. 3.3e-34;
Matches 146; Conservative 111; Mismatches 215; Indels 44; Gaps 11;

OY 3 NSNDNORVIVIGAGLGLSAAISLATAGPSVOLIEKNDKVGKLNIMTKDGTFTDLQPSIL 62
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 ASDADRALVIGSLGGLAAARLGAAGRWVTVIDKLDPVPGRGSSITQEGHFRFDLGPITV 66
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

ESULT 11

ENTRY	KEYWORD	STANDARD	PRT	495 AA.
D	QRTD_RHOSH			
D	Q01671;			
01-JUL-1993	(Rel. 26, Created)			
16-OCT-2001	(Rel. 40, Last sequence update)			
28-FEB-2003	(Rel. 41, Last annotation update)			
ME	Methoxyneurospore dehydrogenase (EC 1.14.99.-).			
CPD.				
N	Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).			
NC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;			
DC	Rhodobacteraceae; Rhodobacter.			
NCBI_TaxID=1063;				
[1]	SEQUENCE FROM N.A.			
NP	STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;			
CC	MEDLINE=92307398; PubMed=1612412;			
CC	Gari E., Toledo J.C., Gibert I., Barbe J.;			
CC	"Nucleotide sequence of the methoxyneurospore dehydrogenase gene			
CC	from Rhodobacter sphaeroides: comparison with other bacterial			
CC	carotenoid dehydrogenases.";			
CC	FEMS Microbiol. Lett. 72:103-108(1992).			
[2]	SEQUENCE FROM N.A.			
NP	STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;			
CC	MEDLINE=95238278; PubMed=7721699;			
CC	Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;			
CC	"Complete DNA sequence, specific in5 insertion map, and gene			
CC	assignment of the carotenoid biosynthesis pathway of Rhodobacter			
CC	sphaeroides.";			
CC	J. Bacteriol. 177:2064-2073(1995).			
CC	-1- FUNCTION: CONVERTS HYDROXYNEUROSPORE TO DEMETHYLSPHEROIDE OR			
CC	METHOXYNEUROSPORE TO SPHEROIDE.			
CC	-1- COFACTOR: FAD (Probable).			
CC	-1- PATHWAY: Carotenoid and chlorophyll biosynthesis.			
CC	-1- SIMILARITY: Belongs to the phytoene dehydrogenase family.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			

```

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; J63204; CAA44886.1; -.
DR EMBL; AJ010302; CAB38743.1; ALT_SEQ.
DR PIR; S23633; S23633.
DR InterPro; IPR002937; Amino oxidase
DR InterPro; IPR008150; Sac phytoene dh.
DR InterPro; IPR002020; NAD BS.
DR Pfam; PF01593; Amino oxidase; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
FT Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 9 42 FAD (ADP PART) (POTENTIAL).
FT CONFLICT 115 115 A -> G (IN REF. 2).
FT CONFLICT 157 157 T -> P (IN REF. 2).
FT CONFLICT 170 170 L -> M (IN REF. 2).
FT CONFLICT 273 273 L -> I (IN REF. 2).
FT CONFLICT 440 443 PHGA -> ATGP (IN REF. 1).
SQ SEQUENCE 495 AA; 52900 MW; 62227931415B253E CRC64;
Query Match 20.8%; Score 541.5; DB 1; Length 495;
Best Local Similarity 29.9%; Pred. No. 9.6e-34;
Matches 151; Conservative 95; Mismatches 22; Indels 37; Gaps 13;
QY 8 RVIVIGAGLGLSAAISLATAGSVQLIEKKDKVGKLNIMTKDGFDDLGPSILNPHI 67
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 7 KVVVVGAGMGLASARLARAGCEVTLLEAPAGPGRVRLTLPVAGPVDAGPTVLTREV 66
QY 68 FEALPTGAGKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETQRRELDKL-GRGTYAQF 126
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 67 FDDIPEVCGKLDHLLTILPQPLARHWLWLDGSTLDTLTDLEANVEAVAAPAGAREAPAF 126
QY 127 ORFLDYKNICTETAGYFAKLDGF-----WDLLKP-YGPLRSLLSFDVF---RS 173
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 127 RRFHDLGARL-----YDAPDRPMWEAARFDLRAIATGALKAPRTWPALLEGMT 174
QY 174 MDQGVRFISDPKLVILNFKYVSSPYDAPALMNLPLYQIYHGLWTVKGGYGMQAQ 233
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 175 LDRLLRLLPFDRLRLQRLRYATYVGGTYGAGVIALI-WAAEARGWALEGGMRLAL 233
QY 234 AMEKLAVELGVEIRLDAVRSIKQDGRACAVKLANGDVLPADIVSGNMEVIPAMEKLLR 293
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 234 ALARLADQGVRLRYGAPVAGILRQGRGTGVQLADGRTLPAADHIVFNGDPAALLAGCLG 293
QY 294 SPASLKKMORFEP-SCSGLVLHLGVDRILYPCLAHNFFYSDHPREFFDAVFKSHRLSD 352
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 294 DGPQDAVPEDRIIPRSLSAWWSYARASGPPLVHNVPFADDPREFFGPI-AAGQMPED 352
QY 353 PTVILVAPCKTDPAPAGACETIKILPHIPHDLPKLTADSYGALRERVLVKLERNGLT 412
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 353 ATLVCIAEDRSQ-GQLPDGPERFEIIMNGP---PGPAPKPEDPACRSRTPDLRQFGLT 408
QY 413 DLRQHIVTEEYWT-PLDIOAKYYSNQGSIYGVVADRFKNLGFKAQPSSELSNLYFVGGS 471
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 409 --FDPVFGETSLTAPSGFASLPASGSIYG-LSPHGALASLKRFLARTALPGLWLAGGG 465
QY 472 VNPGGGMWTLTGSQLVRDKIVADL 496
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 456 AHPGAGVPMALSGREAAAILADL 490
RESULT 12
CRTL_STRCH
ID CRTL_STRCH STANDARD; PRT; 507 AA.
AC P54981; P72447;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTL OR CRTL.

```

Streptomyces griseus.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_taxid=1911;
[1]
SEQUENCE FROM N.A.
STRAIN=JA3933;
MEDLINE=97074881; PubMed=8917308;
Schumann G., Nurnberger H., Sandmann G., Kruegel H.J.;
"Activation and analysis of cryptic crt genes for carotenoid
biosynthesis from Streptomyces griseus";
Mol. Gen. Genet. 252:658-666(1996).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COFACTOR: FAD (probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L37405; AAA91950.1; -;
EMBL; X95596; CAA64850.1; -;
InterPro; IPR000759; Adrxn. reductase.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac phytoene dh.
InterPro; IPR000205; NAD BS.
InterPro; IPR008151; Phyt. dehydro.
Pfam; PF01593; Amino oxidase; 1.
PRINTS; PR00419; ADXRDTASE.
ProDom; PD139017; Phyt. dehydro; 1.
PROSITE; PS00982; PHYTOENE DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SEQUENCE 507 AA; 54509 MW; FBB97F7F6596B2AC CRC64;

Query Match 20.3%; Score 526.5; DB 1; Length 507;
Best Local Similarity 28.6%; Pred. No. 1.4e-32;
Matches 146; Conservative 88; Mismatches 218; Indels 59; Gaps 12;

9 VIVICAGLGGLSAALSLATAGFSVOLIEKNDKVGKLNIMTKDGTDFDLGPSILTMPIHF 68
11 VVWVGAGLAGLAAALHLGAGRRVTVVERDVPGRAGLLESGGFRIDTGTPTVLTMPDLV 70
69 EALFTGAGKMWADYVQIOKVPHRNFFEDGSVIDLCEDAETORRELDKLGPGTYA--- 124
71 EDAFAAVERMADRLELRAPAYARFADGSLDVTGDAEAAVEEFAGARQAVGYR 130
125 -----QFORFLDYSKNLCCTETEAGYFAKGLDGFWDLLKFGP-IRSLSPDVF 171
131 LRILWLERLYRVQMERFD-----TNFDSPLQLAHPDLARLAALGGF 172
172 RMDGVRERFSDPKLVILYFYKYGVSSPYDAPALMULLPYQYHYGLVYVYKGMVGM 231
173 GRLDARIGHFVSDERLRVFSQALYAGVPPARALAAAYAVIAYMDTVAGYFFRGGKHAL 232
232 AQAMEKLAVELGVEIRLDAEYSEIOKQDGRACAVKLANGDVLPAIVVSNNEVIPAMEKL 291
233 PRAMADAASDAGAVLYRGVQRVTRLESGDRVTAV-VTDQEHIPCDAAVLTPD-LFVSYRL 290
292 L-RSPASELXKMQRPSCGLVLHLGVDRILYQLAHNFFYSIDPHREHFDVFKSHRLS 350
291 LGRTFHRPLP-----LRHSFSAVILHTGTDRVPLAAHHTISFGAAWKNTEFELTRTGLM 346
351 DPTTYLVAPCKTDPQAQAPAGCEIILKILPHILD--PDKLLTAEDYSALRERVLVKLER 408
347 SDFSLLITRPTADTDFSLAPPCKHLHYVLAPCNTEVGVREWELELGPYRDELLASLER 406

QY 409 MCLTDLRCHIVTEEVWTPDLQAKYTSNQSGSIYGV--VADRFKNLGFKAPQSGSEL--- 462
DB 407 REMPGLGAAIEBEGLVTPD-----WTAQGHAAAGTSPFSAHTFPQTG--PFRPNLVRG 458
QY 463 -SNLYFVGGSVNPGGGPMWTLSQLVRDKI 492
DB 459 TVNAVLACGCTTPGVGVPTVLISGLAAERI 489

RESULT 13

CRTI_STRSE STANDARD; PRT; 508 AA.
AC P54971;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI.
OS Streptomyces setonii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=38315;
RN [1]_TaxID=38315;
RP SEQUENCE FROM N.A.
RC STRAIN=ISP 5395;
RA Hoshi K.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D55723; BAA09537.1; -;
DR InterPro; IPR000759; Adrxn. reductase.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR000205; NAD BS.
DR InterPro; IPR008151; Phyt. dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR ProDom; PD139017; Phyt. dehydro; 1.
DR PROSITE; PS00982; PHYTOENE DH; 1.
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 508 AA; 54610 MW; 4B6DEF076D31CB5 CRC64;

Query Match 19.8%; Score 514.5; DB 1; Length 508;

Best Local Similarity 28.0%; Pred. No. 1.1e-31;
Matches 143; Conservative 89; Mismatches 221; Indels 57; Gaps 12;

QY 9 VIVICAGLGGLSAALSLATAGFSVOLIEKNDKVGKLNIMTKDGTDFDLGPSILTMPIHF 68
DB 11 VVWVGAGLAGLAAALHLGAGRRVTVVERDVPGRAGLLESGGFRIDTGTPTVLTMPDLV 70
QY 69 EALFTGAGKMWADYVQIOKVPHRNFFEDGSVIDLCEDAETORRELDKLGPGTYA--- 124
DB 71 EDAFAAVERMADRLELRAPAYARFADGSLDVTGDAEAAVEEFAGARQAVGYR 130
QY 125 -----QFORFLDYSKNLCCTETEAGYFAKGLDGFWDLLKFGP-IRSLSPDVF 172
DB 131 LRILWLERLYRVQMERFD-----TNFDSPLQLAHPDLARLAALGGF 173

173 SYDQVRRFTSDPKLVEILNFIKTVGGSPYDAPALMILLPYIOVHYGLVWVKGWYGA 232
 174 RLDAIRGHVSDRLRRVFSQALYAGVTPARALAYAVIAMDVTAGVYPRGGMHALP 233
 233 QAMEKLAVELGVEIRLDAEVSBIQKDGACAVKLANGDVLDPADIVSNMEVIPAMEKLL 292
 234 RAMADAADAGASFRYQSGVTLERSGDRVAV-VTDQERACDAVLTPO-LPVSYRL 291
 293 -RSPASELKMQRFPESCGLVHLGVRLXPQLAHNFFYSDDHREHDFDAVEKSHRLSD 351
 292 GRSPHPLP-----LRHSPSAVILHAGTDTWPNLAHTISFGAWKSTFHELTRIGELMS 347
 352 DPTIYLVAPCKTDPAAQAPAGCEIIKILPHIPELD--PKLLTAEDYSALRERVLVKLERM 409
 348 DPSLLITRPASDPSLAPPSKHLHYVLAPCPNTEVPGVREWRELGPYRDELLAELERR 407
 410 GLTDLRQIHVTEYWTPLDIQAKYVNSQSIYGV---VADRKNLGFKAPORSSEL----- 462
 408 EMPGLGSAIEVEGLVTPVD-----WTAQGHAAAGTPPSVAHTFPQTG---PFRGNLVRGT 459
 463 SNEYFYVGGSVNPGGGMVMTLSGLVRDKI 492
 460 VNAVLAGCGTTPGCVGPTVLISGLAAQRI 489

RESULT 14
 RTD RHOCA
 D CRTD RHOCA STANDARD; PRT; 494 AA.
 C P17059;
 T 01-AUG-1990 (Rel. 15, Created)
 T 01-AUG-1990 (Rel. 15, Last sequence update)
 T 28-FEB-2003 (Rel. 41, Last annotation update)
 E Methoxynurosporene dehydrogenase (EC 1.14.99.-).
 C RTD.
 S Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
 C Rhodobacteraceae; Rhodobacter.
 X NCBI_TaxID=1061;
 N (1)
 K SEQUENCE FROM N.A.
 P STRAIN=SB1003 / St Louis, and BEC404;
 C MEDLINE=89313663; PubMed=2747617;
 X Armstrong G.A., Alberti M., Leach F., Hearst J.E.;
 A "Nucleotide sequence, organization, and nature of the protein
 T products of the carotenoid biosynthesis gene cluster of Rhodobacter
 T capsulatus.";
 T Mol. Gen. Genet. 216:254-268 (1989).
 N (2)
 P SIMILARITY TO CAROTENOID DESATURASES.
 X MEDLINE=90368827; PubMed=2144293;
 X Bartley G.B., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
 A "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
 T crassa are structurally and functionally conserved and contain
 T domains homologous to flavoprotein disulfide oxidoreductases.";
 J Biol. Chem. 265:16020-16024 (1990).
 L L-1- FUNCTION: CONVERTS HYDROXYNUROSPORENE TO DEMETHYLSPHEROIDENE OR
 C METHOXYNUROSPORENE TO SPHEROIDENE.
 C L-1- COFACTOR: FAD (Probable).
 C L-1- PATHWAY: Carotenoid and chlorophyll biosynthesis.
 C L-1- SIMILARITY: Belongs to the phytoene dehydrogenase family.
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 X EMBL; X52291; CAA36537.1; --
 X EMBL; Z11165; CAA77544.1; --
 X PIR; S04406; S04406.
 X InterPro; IPR002937; Amino_oxidase.

DR InterPro; IPR008150; Bac_phytoene_dh.
 DR InterPro; IPR008151; Phytin dehydro.
 DR Pfam; PF01593; Amino_oxidase; 1.
 DR ProDom; PD139017; Phytin dehydro; 1.
 DR ProSite; PS00982; PHYTOENE_DH; 1.
 XW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
 XW Oxidoreductase; FAD; Flavoprotein; NAD.
 FT NP BIND 8 41 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 494 AA; 52312 MW; D1180A023FEB5A9 CRC64;
 Query Match 18.9%; Score 490.5; DB 1; Length 494;
 Best Local Similarity 27.4%; Pred. No. 7.5e-30;
 Matches 141; Conservative 88; Mismatches 226; Indels 59; Gaps 14;
 QY 9 VIVIGAGLGLGSAASLATAFSGVQLIEKDKVGGKLNIMTKDGTFFDGLGPIILTWPHIF 68
 DB 7 VVIGARMGGLAAAI GAAAGLRVTVEAGDAPGKARAVPTGPGPADTGTVLTRHVL 66
 QY 69 EALPTGAGKNMADYVQIQKVEPHNRFNFFEDGSGVIDLCDAETORRELDKL-GPGTYAQFQ 127
 DB 67 DALFAACGTRAEHLTLIPLRLARHFWPDGSSLDLFTDTTEANIEAIRAFAGDKKEAAFR 126
 QY 128 RFLDYSKNLCCTETAGYFAKGLDGFWDLL-----KFGPLRLSLLSPDVFRSMDOGVRR 180
 DB 127 RFDLITLGLWEAFHRSVIAAPKPSDLWRIAAATVTRPQLWPALRPGL-----TMRDLLAH 180
 QY 181 FIDPKLVEILNFIKTVGGSPYDAPALMILLPYIOVHYGLVWVKGWYGAQAMEKLA 240
 DB 181 HFKDPRLAQLFGRYATYVGRPGATPAVLISLWQAEVQ-GVAIRREGMEGVAAALARVAE 239
 QY 241 ELGVEIRLDAEVS EIQKDGACAVKLANGDVLDPADIVVSNMEVIPAMEKLLRSPA-SEL 299
 DB 240 AKGVRFHYGKAKRIVRKEGRVTAIEITGVSI PCGACIFNGDPGALRDGLLGDAARSM 299
 QY 300 KMQRPSPCSGLVHLGVRLXPQLAHNFFYSDDHREHDFDAVEKSHLSDDPTIYLA 359
 DB 300 EKSPFPAPSLSAWVWAFGATPIGVDLAHENVTADPELEFGPI-GAGNPREPIYI-- 356
 QY 360 PKCTD-PAQAPAGCEIIKILPHIPLDPKLLAT-----AEDYSALRERVLVKLERM 409
 DB 357 -CAQDREMQAP-----VPEIERFEIIMNGPAGHQPFPEAEACQKARTFPMLAAM 404
 QY 410 GLT---DLRQIHVTEYWTPLDIQAKYVNSQSIY-----GVVADRFLKLGPKAPORSSE 461
 DB 405 GLTFSPDPETRALT-----TPALLSRFPGLSGIAIYGGSPGTLAT-----PRLPLARTG 454
 QY 462 LSNLYFVGGSVNPGGGMVMTLSGLVRDKIVAD 495
 DB 455 LKGLYLAGGTHFGAGVPAVLTSGTEAARALLAD 488

RESULT 15
 CRTI SYNY3
 ID CRTI SYNY3 STANDARD; PRT; 472 AA.
 AC P23273;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
 GN PDS OR CRTD OR SLR1254.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 N (1)
 K SEQUENCE FROM N.A.
 P MEDLINE=92256820; PubMed=1581575;
 X Martinez-Perez I.M., Vioque A.;
 A "Nucleotide sequence of the phytoene desaturase gene from
 T Synechocystis sp. PCC 6803 and characterization of a new mutation
 T which confers resistance to the herbicide norflurazon.";
 T Plant Mol. Biol. 18:981-983 (1992).
 N (2)
 P REVISIONS TO C-TERMINUS.

A Vioque A.;
L Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
P [3]

SEQUENCE FROM N.A.

K MEDLINE=97061201; PubMed=8905231;
A Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
A Miyajima N., Hikosawa M., Sugitara M., Saito S., Kimura T.,
A Hosouchi T., Hara A., Muraki A., Nakazaki N., Naro K.,
A Okumura S., Shimizu S., Takeuchi C., Wada T., Watanabe A.,
A Yamada M., Yasuda M., Tabata S.;

"Sequence analysis of the genome of the unicellular cyanobacterium
Synchocystis sp. strain PC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
DNA RES. 3:109-136(1996).

1-! FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.

1-! COFACTOR: NAD, NADP, or FAD (Probable)

1-! ENZYME REGULATION: Inhibited by the herbicide norflurazon in a
non-competitive way.

1-! PATHWAY: Carotenoid biosynthesis.

1-! SUBCELLULAR LOCATION: Membrane-associated (Probable).

1-! SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

EMBL; X62574; CAA44452.1; -

EMBL; D90909; BAA17847.1; -

PIR; S74886; S74886.

InterPro; IPR007059; Adrindx reductase.

InterPro; IPR001613; Amineoxid fl.

InterPro; IPR002937; Amino oxidase.

Pfam; PF01593; Amino oxidase; 1.

PRINTS; PR00419; ADXRDASE.

PRINTS; PR00757; AMINEOXDASEF.

M Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;

M Membrane; Herbicide resistance; Complete proteome.

I NP BIND 7 23 FAD (ADP PART) (POTENTIAL).

I VARIANT 195 195 R -> C (CONFERS RESISTANCE TO THE

HERBICIDE NORFLURAZON).

I SEQUENCE 472 AA; 52920 MW; 81D089A6DAA28758 CRC64;

Query Match 6.7%; Score 174.5; DB 1; Length 472;

Best Local Similarity 22.9%; Pred. No. 9e-06;

Matches 129; Conservative 65; Mismatches 186; Indels 183; Gaps 31;

Y 8 RVIVAGIGLGLSALISLATAGSVQLEKNDKVGKGLNI-MTKDGFYFDLGPILTMHP 66

b 2 RVVIAGAGLAGLACAKYLADAGFTPVLLRRDVLGGKIAAKWDEGDWYETGLHI----- 56

Y 67 IFEALFTGAGKQADYVQIQKEP--HWRNPFDCGSVIDLCEDAETQRRELDKLPQTYA 124

b 57 -----PFGAYPNMLQFKELDIEDRLQW---EHSNIFNQPEK-----PGTYS 96

Y 125 QFORFLDY-----SKNLCETEAGYFAKGL-----DGF-WD--L 155

b 97 RFD-PFDIPAPINGLVAILLRNDMLTWPKEKIFGLGLPAIVQGSVVEENDKRTWSEWM 155

Y 156 LKPYGLRLLSFDVFRSMDQVRFSIDPK-----LVEILNYFIKVGSSPY---DA 205

b 156 AKQNIIPR--IKKVFIAKSKAL-NFI-DPDEISATILLTALNRLFKVNGSKMAFLDGA 211

Y 206 PALNNLLPYIQHYGLWYVKGMYGMAQAMEKLAVELGVRLDAVSEI-QKQDGRACA 264

b 212 PPERLCQPLVDY-----ITERGGEVHINKPLKLEILLNEDGSKVG 250

Y 265 VKL-----ANGDVLPAIVVSNNEVIPAMEKLLRSP-----ASELKQKQRFPSCSGLVLHL 316

Db 251 YLIRGLDGAPDEVITADLYVSAMPVDP-LKTWVPAPWREYEPFKQIOGLE-GVPVINLHL 308

QY 317 GVDRLYPOLAHENF-----FYSDFP---REHPDAVFKSHRLSDDPITYLVAPCKTDP 366

Db 309 WFDRLKLTIDHLLFSRSPILLSVYADMSNTCREYSD-----PKSMLELVLA-----PA 356

QY 367 Q-----APAGCEIILKP-HIHLDPDKLLTAEDYSALKRERVVLKLRMGLTDL 414

Db 357 QDWIGKSDDEEIVATWATWATIKQLFPQHFGNDNPARL-----L 392

QY 415 RGHVITSEYTWTPDLDTQAKYYSNCGSIYGVVADRFKNLGFKAPQSSSELSNLYFVCGSYNP 474

Db 393 KSHVVK-----TPRSVYKATPGRQAC-----RPDQKTSVPNFYLAGDFTWQ 433

QY 475 G--GGMPWVTLSSQLVRDKIVAD 495

Db 434 KYLGSMEGAVLSGKCAQAIAAD 456

Search completed: February 29, 2004, 14:45:02

Job time : 11.3764 secs


```
;Accession: A39273
;Status: preliminary
;Molecule type: DNA
;Residues: 1-492 <ARA>
;Cross-references: GB:M38423; NID:g148401; PIDN:AAA24820.1; PID:g148402
;Note: the authors translated the codon CAG for residue 181 as Phe, TCC for residue 187
;Note: strain Eho10; ATCC 39368
;Genetics:
;Gene: crtI
;Superfamily: phytoene dehydrogenase
;Keywords: carotenoid biosynthesis; oxidoreductase

Query Match      29.6%; Score 770; DB 2; Length 492;
Best Local Similarity 34.7%; Pred. No. 6.1e-52;
Matches 174; Conservative 93; Mismatches 213; Indels 22; Gaps 6;

Y 7 QRVIVIGAGLGLSAAISLATAGSFVQLIEKNDKVGKLNIMTKDGFTEFDLGPISILTMPIHFEA 66
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 2 KXTWVIGAGCGGLAALRAIRLQAAGIPVLLERQDKPGRAYVWHQDGFTEFDAGPTVITDPT 61

/ 67 IFEALFTGAGKNMADYVQIQKVEPHWRNFFEDSGVIDLCEAETQRRELKLGPGTYAQP 126
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 62 ALEALFTLAGRMEDYVRLPKPYRLCWESGKTLDYANDSAELEAQITQFNPRDVEGY 121

/ 127 QRFLYSKNLCTETETAGYPAKG---LDGFWDLLKFGPLRLSILGFDVPRSDQGVRRFIS 183
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 122 RFLAYSQAVFCG---GYLRGSPVPLSPFQWLR-AGP-QLLKLAQWSYQSVSRFIE 175

/ 184 DPKLVEILNYFYKYGSSPYDAPALMNLPIYQHYGLVYKGMGMAQAMEKLAVELG 243
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 176 DEHLQAQFSFSLVGGNPFPTSSITVTLHALEREWGVFPFGGTGALVNGVMKLFDTLG 235

/ 244 VEIRLDAEYSEIKQDGRACAVKANGDVLPAIDIVSNMVIPIAMEKLL-----RSP 295
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 236 GRIELNARVEELVADNRVSVQVRLADGRIFDTDAVASNADVNTYKLLGHHPVQKRAA 295

/ 296 ASBLKMQRFEPSCSGLVHLGVDRLYPQLAHNPFYSDHPRHFDVAFKSHRLSDDPTI 355
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 296 ALERKSM-----SNSLFVLYGLNGPHQSALHTTCFQPRYRELIDIFTGSALADDFSL 350

Y 356 YLVAPCKTDPAAQAGCEIKILPHIPLDPDKLLTAEDYSALRERVLVLERMGLTDLR 415
D 351 YLHSPCVTDPSLAPGCSFYVLAPVPHLGNAPLDWAQEGPKLDRIPDYLEERYMGLR 410

Y 416 QHVTVEYWTPLDIQAKYNSQGSYGVVADRPNLGFAPQSRSELSNLYFVGGSVNP 475
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
D 411 SGLVTRIFTPADFTDLDAHLGSAFSIEPLLTSQAWFRPHNRDSDIANLYLVGAGTHRG 470

Y 476 GGMPMYLLSQGLVRDKIVADLQ 497
D 471 AGIPGVVASAKATASLMIBDLQ 492

RESULT 5
52586
Tytoene dehydrogenase (EC 1.3.-.-) - Erwinia herbicola
;Species: Erwinia herbicola
;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999
;Accession: S32586
;Lin, Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.
;Jl. Gen. Genet. 245, 417-423, 1994
;Title: Transcriptional activation of flanking sequences by Tn1000 insertion.
;Reference number: S52583; MUID:95107237; PMID:7808390
;Accession: S52586
;Status: preliminary; nucleic acid sequence not shown; translation not shown
;Molecule type: DNA
;Residues: 1-492 <NHA>
;Cross-references: EMBL:M90698; NID:g148393; PIDN:AAA21263.1; PID:g148397
;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
;Superfamily: phytoene dehydrogenase
;Keywords: oxidoreductase

Query Match      26.4%; Score 685; DB 2; Length 492;
Best Local Similarity 31.5%; Pred. No. 2.6e-44;
Matches 162; Conservative 120; Mismatches 191; Indels 42; Gaps 14;

QY 8 RVIVIGAGLGLSAAISLATAGSFVQLIEKNDKVGKLNIMTKDGFTEFDLGPISILTMPIH 67
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 2 KTVIGAGVTGLAAARATASQGHVTFERKNNVGVGRMQLKQDGFTEFMGFTIWMMPDV 61

QY 68 FEALFTGAGKNMADYVQIQKVEPHWRNFFEDSGVIDLCEAETQRRELKLGPGTYAQP 127
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 62 YKDFVTAQCKNYEDIELRQLRYIYDVFDHDDRTVPTDLAELQGMLESIEPGSTHGM 121

Best Local Similarity 33.3%; Pred. No. 2.5e-45;
Matches 164; Conservative 90; Mismatches 226; Indels 12; Gaps 5;

QY 11 VTGAGLGLSAAISLATAGSFVQLIEKNDKVGKLNIMTKDGFTEFDLGPISILTMPIHFEA 70
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 6 VTGAGFGGLAALRAIRLQAAGIPVLLERQDKPGRAYVVEDQGFTEFDAGPTVITDSALIEE 65

QY 71 LFTGAGKNMADYVQIQKVEPHWRNFFEDSGVIDLCEAETQRRELKLGPGTYAQPRL 130
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 66 LFTLAGLQKDYVELLPVAPFYRLCWESGKVFYNDQALEAQIQFNPDRDVEGYRFL 125

QY 131 DYSKNLCTETETAGYPAKGGLDGFWDLLKFGPLRS---LLSPDFVFSMDQGVRRFISDPKL 187
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 126 DTSRAVFKG---GYLRGTVPP---LSFRDMLRAAPOLAKLOAMTVTSKVASYTEDEHL 179

QY 188 VEILNYFYKYGSSPYDAPALMNLPIYQHYGLVYKGMGMAQAMEKLAVELGVEIR 247
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 180 RQAFSPSLLVGGNPFATSSITVTLHALEREWGVFPFGGTGALVKGMIKLPDLGGGEV 239

QY 248 LDAEYSEIKQDGRACAVKANGDVLPAIDIVSNMVIPIAMEKLL-RSPAS--ELKMQOR 304
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 240 LNAKVSHEMTTGGDTTETAEVHLEDGRFPFTRAVASNADVVHTYRDLLSQHPAAVKQSKLQT 299

QY 305 FEPSCSGLVHLGVDRLYPQLAHNPFYSDHPRHFDVAFKSHRLSDDPTIYLVAPCKTD 364
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 300 KRMSNLSFVLYGLNHHHDLQAHTVCGPRYRELHIEIFNHDGLADDFSLYHAPCVTD 359

QY 365 PAQAPAGCEIIKILPHIPLDPDKLLTAEDYSALRERVLVLERMGLTDLRQHIVTSEY 424
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 360 SSLAPEGCSYVTLAPVPHLGTANTJOWTEGPRLDRIFEYLEQHYNPEGLRSQLVTRMF 419

QY 425 TPLDIOAKYNSQGSYGVVADRPNLGFAPQSRSELSNLYFVGGSVNP 484
DB 420 TFPDPRDLNAYGSAFSEVPVLTQSAWFRPHNRDKTINNLYLVGAGTHPGAGIGVIGS 479

QY 485 GQLVRDKIVADL 496
DB 480 AKATAGLMEIDL 491

RESULT 6
B90061
Squalene synthase [imported] - Staphylococcus aureus (strain N315)
;Species: Staphylococcus aureus
;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 02-Aug-2002
;Accession: B90061
;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: B90061
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-502 <KUR>
A;Cross-references: GB:BA000018; PID:g13702511; PIDN:BA843652.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: crtN
C;Superfamily: phytoene dehydrogenase

Query Match      25.9%; Score 672; DB 2; Length 502;
Best Local Similarity 31.5%; Pred. No. 2.6e-44;
Matches 162; Conservative 120; Mismatches 191; Indels 42; Gaps 14;

QY 8 RVIVIGAGLGLSAAISLATAGSFVQLIEKNDKVGKLNIMTKDGFTEFDLGPISILTMPIH 67
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 2 KTVIGAGVTGLAAARATASQGHVTFERKNNVGVGRMQLKQDGFTEFMGFTIWMMPDV 61

QY 68 FEALFTGAGKNMADYVQIQKVEPHWRNFFEDSGVIDLCEAETQRRELKLGPGTYAQP 127
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 62 YKDFVTAQCKNYEDIELRQLRYIYDVFDHDDRTVPTDLAELQGMLESIEPGSTHGM 121
```

128 RFLD--YSKXNLCCTETESGYF-----AKGLDGFWDLLKTYGPRLSLLSFVDFRSM---DOG 177
122 SFLTDVYK---YFIARYPLERTYRKPSD-----FYN-MTSLVQGGAKLTLNHAQOL 170
178 VRFPISDPKVEILNFIKVVGSPPDAPALMNLPIQVHGLWTVKGMVGMQAQAMEK 237
171 IEHYIDNEIKQLLAFCGLYIGDPKGSLSYIIPIWEMFGVHFIKGMVGMQAQGLAQ 230
238 LAVELGVEIRLDAVSE--IQODGRACAVKLANGDVLFPADIVVSNMEVIPAMEKLLRSP 295
231 LNKDLGVNIEBNAEIEGIIIDPKFTRADAIV--NGDIRKFCKILCTADFPFSAVESLMPDF 289
296 ASELK----KMQRFEPSCSGLVHLGVD--RLYPQLAHNFFYSDHPRHEDFAVEKSHRLS 350
290 APIKYPFKIALDLVSCAFLEWIGIDIDVTQVRLHNVISDDFRGNIEIFPG--RLS 348
351 DDPTIYLVAECTDPQAQAPACGEIILKILPHIPLD-----PKLLTAEDYSALRERV 403
349 YDPSIYVYFAVADKSLAPEGKTGIYVLPTEPKTGSGIDWSDEALTQQ-----IKIET 404
404 VKLERMGL--TDLRQHIVTEYWTPLDIAQKYNSQGSIVGVVADRPKNLGFKAPOSSEL 462
405 RKLATIEVFEDIKSHIVSEITFTPNDFPEQYHAKFGSAFGLMPTLAQSNYIRPQNVSRDY 464
463 SNLYFVGSVNPFGGMPVMTLSQGLVRDKIVADLQ 497
465 KDLYFAGATHPGAGVPIVLTSAKITVDEMIKDIE 499
RESULT 7
137802
phytoene dehydrogenase (EC 1.3.-.-) crtI - Erwinia uredovora
Species: Erwinia uredovora
Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 20-Jun-2000
Accession: D37802
Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim
i. Bacteriol. 172, 6704-6712, 1990
Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio
Reference number: A37802; MUID:91072214; PMID:2254247
Accession: D37802
Status: preliminary
Molecule type: DNA
Residues: 1-492 <MIS>
Cross-references: GB:D90087; NID:G216681; PIDN:BAAL4127.1; PID:G216685
Superfamily: phytoene dehydrogenase
Keywords: oxidoreductase
Query Match 25.8%; Score 671; DB 2; Length 492;
Best Local Similarity 32.7%; Pred. No. 3e-44;
Matches 161; Conservative 89; Mismatches 230; Indels 12; Gaps 5;
11 VIGAGLGLSAAISLAPAGFSVOLIENKDKVGGKINIMTKDGTFFDLGSPSILTMPHIFEA 70
6 VIGAGFGGLALALRQAAGIPVLLLEQDPKGRVAVYEDQGTFFDAGTPTVITDPSAIEE 65
71 LFTGAGKNADYVIOKVPEHWNFRFEDGSVIDLCDAETQRELDKLGPGITVAOPREL 130
66 LFALAGKQKEYVELLPVTFYRLCWESGKVFDYDNDQTLKLEAQIQFNPDRVEGTRQL 125
131 DYSKNLCCTEAGYFAGLGDGFDLLKXYGLRS---LLSFDVFRSMQGVRRFISDPKL 187
126 DYSAVFEK---GYLKLGTVPF---LSFDMRLRAAPQLAKLQAWRSVYSKVASIYEDHL 179
188 VEILNFIKVVGSPPDAPALMNLPIQVHGLWTVKGMVGMQAQAMEKLAIVELGVEIR 247
180 RQAFSPHLLVGGNPFATSIYTLIHALEREWGVNFRFGTGALVQGMKILFQDLGGEVV 239
248 LDAVSEIQODGRACAVKLANGDVLFPADIVVSNMEVIPAMEKLL--RSPAS--ELKQMQR 304
240 LNAVSEHVEITGNKIEAVHEDGRFRFTQAVASNADVWHYTRDLSQHPAAVQSKNLQT 299
305 FEPSCSGLVHLGVDRLYPLQAHNFFYSDHPRHEDFAVEKSHRLSDDPTIYLVAECKTD 364

300 KRMSNSLFIYFGLNHHHDLAHTVCFGRPRZELIDEIFNHDGLAEDFSLYLHAFCTVD 359
365 PAQAPACGEIILKILPHIPLDPPKLLTASDYSALRERVLVKLERMGLTDLRQHIVTEETN 424
360 SSLAPEGCGSYVYLAFLVPHLGTANLDMWTVEGPKLRDRIIFAYLEQHYMPGLRISQVLRMF 419
425 TPLDIAQKYNSQGSIVGVVADRFKILGFKAPOSSELSNLYFVGSVNPFGGMPVMTLS 484
420 TPDFRQLNAYGSAFVSPVLTQSAWFRPHNRDKTITNLVVGAGTHPGAGIPGVIGS 479
485 GQLVRDKIVADL 496
480 AKATAGLMLLEDL 491
RESULT 8
H84320
phytoene dehydrogenase [imported] - Halobacterium sp. NRC-1
Species: Halobacterium sp. NRC-1
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C/Accession: H84320
R/NG:W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, I.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; L
A/Title: Genome sequence of Halobacterium species NRC-1.
A/Reference number: A84160; MUID:20504483; PMID:11016950
Accession: H84320
Status: preliminary
Molecule type: DNA
Residues: 1-512 <STO>
Cross-references: GB:AE004437; NID:G10581152; PIDN:AAG19932.1; GSPDB:GN00138
Genetics:
Gene: crtII
Query Match 25.5%; Score 662.5; DB 2; Length 512;
Best Local Similarity 32.4%; Pred. No. 1.5e-43;
Matches 169; Conservative 81; Mismatches 217; Indels 55; Gaps 11;
9 VIVIGAGLGLSAAISLAPAGFSVOLIENKDKVGGKINIMTKDGTFFDLGSPSILTMPHIF 68
9 VAVIGAGFGGLSTACYLADAGADVTVVEKTDQIGRASTLERDGRFDMGSPSYLMPDVF 68
69 BALFTGAGKNADYVIOKVPEHWRNPFEE-----DGSVIDLCEDAET 110
69 EEFPAFSDSPSYTYGLTRLDPHYRIFPKDNEGRFRGDPAGLNVDTGDTIDVTDRSQ 128
111 QRRELDKLGPGTYAQFORFLDYSKNLCTETETAGYFAGLGDG-----WDLKFTY 160
129 VKQYFDAYEPGAGVDLDDYLAQAE-----NYEVGMEHFVKTRPRVRDMDPKLAE 180
161 PLRSLLSFDVFRSMQGVRRFISDPKLVEILNFIKVVGSPPDAPALMNLPIQVHYG 220
181 YARGL---TLTGSNQDHVEQYFDPKQLQVMOVTLVFLGSPDPTTALYNLMSHVDFGLG 237
221 LWYVKGGMVGMQAQAMEKLAIVELGVEIRLDAVSEIQKQGRACAVKLAN--GDVLPAIV 278
238 VYTEGEGGVGVGIGITALARELGVFRTGHPVSAIK---GRGCGFKLDTPEGDAVLADV 294
279 VSNMEVIPAMEKLL----RSPASELKKMQRFEPSCSGLVHLGVDRLYPLQAHNFFYSD 334
295 VSDADYATHEQALLPAQKQYDADYWDARTYAP--SAFLLYLGVGEGDVEPLAHTLVLPS 352
335 HPRHEDFAVEKSHRLSDDPTIYLVAECKTDPAQAPAG--CHEIKILPHIPLDPPDKLLTAE 393
353 DMWTHFAQIIFDDAMPEDDPAYILCVFSKTDVTPVDPGHNSLPAVFPVAGLD--DTPAVRE 411
394 DYSALRERVLVKLERMGLTDLRQHIVTEYWTPLDIAQKYNSQGSIVGVVADRPKNLQF 453
412 SY---RDLVDDIAENTGVLDRLKIIVVEERFSVSEPADRYNSHQGTALGLAHTLRQALF 468

Accession: S43139
Status: preliminary
Molecule type: DNA
Residues: 1-583 <RUI>
Cross-references: EMBL:X78434; NID:g468556; PIDN:CAA55197.1; PID:g468557
Genetics:
Introns: 197/2
Superfamily: phytoene dehydrogenase
Keywords: oxidoreductase

Query Match 24.7%; Score 641.5; DB 2; Length 583;
Best Local Similarity 31.3%; Pred. No. 7.6e-42;
Matches 157; Conservative 100; Mismatches 223; Indels 21; Gaps 11;

7 QRVITGAGLGGISAAISLATAGFSVOLIEKNDKVGKLNIMTKDGFDFDLGPSILTMPH 66
b 5 KHVITGAGGAGGATTAARLAREGIKTVVVEKNFNGGRCRLNHNHGFDDQGPSILYMPK 64
67 IFEALFTGAGKNMADYVQIOKVEPHWRNFFEDGSVIDLCEDAETORRELKLGPGTYAQ 125
b 65 LFEAFEALEDEKIEDHVELLRCHNNYKVHFDGDKIQISSLSDLSRMKPEMERIEGPD--G 121
126 FQRFIDYKNLCTETAGY---PAKGLDGFMDLKP-YGFLSLLSFDVFRSMDQGVRRF 181
b 122 FLRFIDFMKESHYEGGVEMAIKQNFETIKWLRLQYVP--ALFLHIFDFVYRAAKY 179
182 ISDPKLVIELNYFIKYVSSPYDAPALMNLPIQYHGLWYKGMYGMAQAMEKLAVE 241
b 180 PKTKGEMAFITQSMYKMSPYDSPAVYNLLQYTEPAEGIWKYFGGFTVIOKLENIATE 239
242 LGVEIRLDEVESEIQOD--GRACAVKLANGVLDADIVVSNMEVIPAMEKLRLSPASE 298
b 240 KFGARFIYEAPAKINTDRKGXVGTGLQSGVEIADAVCNADLVYAHNLLPFCRWT 299
299 LKMKORPEPCSGLVHLGVDRLYPOLAHNFFYSHPREHFDVAFKSHRLSDPTIYLV 358
b 300 TNLAEKLTSSSISFVWSLKRVPVPELDVNIFLAFAFKESFDEIFTDKMPSLSFYVN 359
359 APKTDPAQAPAGCEIILKILPHLPDCKLLTAEDYSAL-----RERVLVKLE-RMGLTLD 413
b 360 LPSRIDPTAAPPGKDSNIVLPIGEM-KSKTNEADYTWIKRARKVLEVLRLGLTN 418
414 LRQHIVTEEYMTLDIOAKYVSNQSGSYGVVADRFRKMLGFKAPQRS--ELSNLYFVGS 471
b 419 FIDIVEHEEVNDPSIWOKKFNLRGSLGLSHDVLQVLFWRFPSTQDSTGRYKNLFPVGNAS 478
472 VNPGGGMPMTLSQLVRDKI 492
b 479 THPGTGVPIVLGSKLTSDQV 499

RESULT 12
33306
phytoene dehydrogenase (EC 1.3.-.-) - Myxococcus xanthus
A:Alternate names: phytoene desaturase
S:Species: Myxococcus xanthus
Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 29-Sep-1999
Accession: S35306; S27594
Fontes, M.; Ruiz-Varquez, R.; Murillo, F.J.
MBO J. 12, 1265-1275, 1993
Title: Growth phase dependence of the activation of a bacterial gene for carotenoid synthesis
Reference number: S35306; MUID:93223667; PMID:8467787
Accession: S35306
Molecule type: DNA
Residues: 1-529 <FOX>
Cross-references: EMBL:M94727; NID:g150079; PIDN:AAA25390.1; PID:g150080
Genetics:
Gene: CarC
Superfamily: phytoene dehydrogenase
Keywords: carotenoid biosynthesis; oxidoreductase
Query Match 24.6%; Score 638; DB 2; Length 529;
Best Local Similarity 30.3%; Pred. No. 1.2e-41;

Matches 155; Conservative 110; Mismatches 208; Indels 38; Gaps 10;
Qy 7 QRVITGAGLGGISAAISLATAGFSVOLIEKNDKVGKLNIMTKDGFDFDLGPSILTMPH 63
Db 9 KHVITGAGGAGGATTAARLAREGIKTVVVEKNFNGGRCRLNHNHGFDDQGPSILYMPK 68
Qy 64 MPHIFEALFTGAGKNMADYVQIOKVEPHWRNFFEDGSVIDLCEDAETORRELKLGPGTY 123
Db 69 LPOVLEQIFERRAARLEEDYVKLLPLDVTNTRVHFMDGTDLDTTRHLDRMEAEALAKFGRQA 128
Qy 124 AQQRFLD-----YSKNLCTETAGYFAKGLDGFMDLKFYGP--LRSLSPDFVR 172
Db 129 SAUKQWDEGREKIGIAYOKFICTSA-----DNLGYAPWELAPTLEKPKWQ 175
Qy 173 SMDQGVRRFISDPKLVIELNYFIKYVSSPYDAPALMNLPIQYHGLWYKGMYGMA 232
Db 176 TLYRQLDGFDDHDDRVYALAYPSKYLGLHPTTCSSVFSVIFLELAFGVHVEGGFRSL 235
Qy 233 QAMEKLAVELGVEIRLDAEVESEIQODGRACAVKLANGVLDADIVVSNMEVIPAMEKL 292
Db 236 RGMWRCARDIGATFRGTGTPVKRVVDAGRAVGLVGGEVLDADAVVNVADLAYAARSLL 295
Qy 293 RSPASELK-----MORFEPSCSGLVHLGVDRLYPOLAHNFFYSHPRE-HFDVAFKS 346
Db 296 PAEABEGSLTDAALERAKYSCSTFWAYVGLDVTYADLPHLIYLSESAARRTDRDALEDR 355
Qy 347 HRLSDPTIYLVAPCKTDPQAOPAGCEIILKILPHIPHL-D-PDKLLTAEDYSALREERVLVK 405
Db 356 HVDLEDPPFYVCPNPGVTPSPGAPAGASTLYLVPTNTGRPVQDWKTE--QALREIIPAM 413
Qy 406 LERMGTLDRLOHIVTEBYMTPLDIOAKYVSNQSGSYGVVADRFRKMLGFKAPQ-RSSELSN 464
Db 414 LEKVGLGVREHREERYFTAEWTWRDDFNVFRAVFN-LSHTWLQGLPLRPVKYKRRDIEG 472
Qy 465 LYFVGGSVNPGGMPMTLSGQLVRDKIVAD 495
Db 473 LYFVGGGTHRPGSGLLTIMESANTAADYLTRE 503

RESULT 13

48646
phytoene dehydrogenase (EC 1.-.-.-) [validated] - Cercospora nicotianae
S:Species: Cercospora nicotianae
Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
Accession: T48646
Ehrenshaft, M.; Daub, M.E.
Appl. Environ. Microbiol. 60, 2766-2771, 1994
Title: Isolation, sequence, and characterization of the Cercospora nicotianae phytoene
Reference number: Z24498; MUID:94368091; PMID:8085820
Accession: T48646
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-621 <EHR>
Cross-references: EMBL:U03903; NID:g433144; PIDN:AA86988.1; PID:g433145
Experimental source: ATCC 18366
Genetics:
Gene: PDH1
Function: <PDH1>
Function: EC 1.-.-.- [validated, MUID:94368091] phytoene dehydrogenase
Function: <PDH2>
Description: dehydrogenation of phytoene
Pathway: carotenoid biosynthesis
Superfamily: phytoene dehydrogenase
Keywords: oxidoreductase

Query Match 24.3%; Score 631.5; DB 2; Length 621;
Best Local Similarity 29.0%; Pred. No. 5e-41;
Matches 162; Conservative 105; Mismatches 214; Indels 77; Gaps 12;

Qy 10 IVIGAGLGGISAAISLATAGFSVOLIEKNDKVGKLNIMTKDGFDFDLGPSILTMPH 69
Db 11 IVIGVGGVSTAAARLARGFHVTVLEKKNPTGRCSLIHHEGTRFDQGPSLLLPGLFH 70

70 ALFTGAGKMM-ADYVQIQVBPWHNFEDSGVIDLCEAETORRELDKL-GPGTYAQFQ 127
71 RTFAELGSLGEGYKLLKCPNMYIHSEDEKTKTSLSDLSVMKTEVKEGEGYRIL 130
128 RFLQYKNCLETETEAGYFAKGLDGDWDLKFKYGP--LRSLSPDFVFRSMOQGVRRFISDP 185
131 EPLKESHGHYELSVREVLNRNFEGLTAMLR---PEFLRHLLQLHPFESIWRACKYFWE 187
186 KIVEILNFIKYVGSSPYDAPALMNLPPYQVHYGLWTVKGMYGMAQAKLAVELGVE 245
188 RLRRVFTTGSYMGMSPPDAGTYSLLQYTLASGIWYVPGGFRVVEALVKIGEREVD 247
246 IRLDAEVSII--QKQDGRACAVKLANGDVLDPADIVVSNMVEVIPAMEKL--RSPASELKX 301
248 FNETAVKXKLLSBDGVAKGVELEDGRLEADVVVWNSDLVAYAYEKLLPIKTPYAESLK 307
302 MQRFPSCGLVHLGVDRLYPOLAHNFFSYDHPREHFDVFKSHRLSDPTIYVAPC 361
308 GR--PGSCSSISFFWALDRQVPELAHNFILADEYRHSFDSIPKGLHLPDEPSFYVWPS 365
362 KTDPAQAPAGCEIILKILPHIL-----DPDKL-LT---A 392
366 RYDSTAAPGKDSVVVLVPGVHLLBEDRHSQAQHSASRNHGISSASPPDQPCGLTPTK 425
393 EYDSAL----RRVLVKLERKGLTDLRQHIWTEYWPDLQAKYYNSQGSIVGVADRF 448
426 QWPAMISLARKTILSTIQSRNVDLTLIHESTNSPLSKWQTFNLDRAIGLGLSHSFF 485
449 KNLGFKAPQRS-----SELSNLYFVFGSVNFGGG 477
486 NVLCFRPTTRARKPGAFDQLLKFGVLGRAEVIIDAFGRGKDIKGLYMWGASAHFGTG 545
478 MNVTLSQLVRDKIVAD 495
546 VPICLAGGALVAEQICGD 563

RESULT 14
35919
Species: Neurospora crassa
Date: 09-Nov-1990 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
Accession: A35919
Schmidhauser, T.J.; Lauter, F.R.; Russo, V.E.A.; Yanofsky, C.
Title: Cloning, sequence, and photoregulation of al-1, a carotenoid biosynthetic gene
Reference number: A35919; MUID: 90377195; PMID: 2144609
Accession: A35919
Status: preliminary
Molecule type: DNA
Residues: 1-595 <SCH>
Cross-references: GB:M57465; GB:M33867; NID:g168746; PIDN:AAA33555.1; PID:g168747
Note: the authors translated the codon CTC for residue 157 as Glu, and GAG for residue 158
Superfamily: phytoene dehydrogenase
Keywords: transmembrane protein

Query Match 24.0%; Score 624.5; DB 2; Length 595;
Best Local Similarity 29.7%; Pred No. 1.6e-40;
Matches 158; Conservative 110; Mismatches 201; Indels 53; Gaps 17;

5 DNQR---VIVIGAGLGSLSAISLATAGFSVQLIEKNDKVGKLNIM-TKDGTFDILGDS 60
3 ETQRPRAIIVGAGAGGIAVAARLAKAGVDVTILEKNDFTGRCGLIHTKAGYRFDQGS 62
61 ILTMPIFEALFTGAGKMM-ADYVQIQVBPWHNFEDSGVIDLCEAETORRELDKL- 118
63 LLLLPLGPRETFEDLCTLEQEDVELLOCFPNYNIWFSDDGRFPTDNATMKVEIKWE 122
119 GPGTYAQFQFIDYKSNLCTETETAGYFAKGLD-----GFWDLLKFKYGP----LR 163
123 GPD---GFRVYL-----SWLAEGHQHYETSRLHVLHNRNFKSILELADPLRVVT 167
164 SLLSFDVFRSMOQGVRRFISDPKILVEILNFIKYVGSSPYDAPALMNLPPYQVHYGLWY 223

Db 168 LLWALHPFESIWHRAGRYFKTDRMQVTFATMTMGSPFPDAPATFYSILQYSELAEGTW 227
QY 224 VKGMYGVAQAAMEKLAVELGVEIRLDAEVSIIQKDG-----RACAVKLANGDVLPA 277
Db 228 PRGGPHKVLDAVLKIGERMGVYRLUNTQSVQLTDGGKNGKPKATGVQLENGEVLNADL 287
QY 278 VVSNNMEVIPAMEKLARSASELK---QMRFPSCGLVHLGVDRLYPOLAHNFPYS 333
Db 288 VVVDNLVVTYNNLLPKBIGGKKYANKLNKRKASCSSISFYWSLSGMAKELETHNIFLA 347
QY 334 DHPRHFDVAVFKSHRLSDPTIYVAPCKTDPAPACACE-IILKLP--HI-----PHLD 385
Db 348 EYKESFPAIFERQALPDPSFYIHVPSRVDPSPAPPORDAVIALVPVGHLLQNGQPSLD 407
QY 386 PKLKTABEDYSALRERVLVKLE-RMGLTDLRQHIWTEYWPDLQAKYYNSQGSIVGV 443
Db 408 WPTLV-----SKARAGVLATIQARTGLS--LSPLITEIVNTPYTWETKENSXGAILGL 460
QY 444 VADRPKNLGFAPKAPORSSELSNLYFVFGSVNFGGMPMTLSGQLVRDKIVAD 495
Db 461 AHDFFNLAFRPTTKAQGMNAYFVGASTHEFTGTVPIVLAGAKITAEQILRE 512

RESULT 15
T50910
Phytoene dehydrogenase [imported] - Rubrivivax gelatinosus
C;Species: Rubrivivax gelatinosus
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
Accession: T50910
R.Nagashima, K.V.; Igaraishi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynt
A;Reference number: Z52570
A;Accession: T50910
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-511 <NAG>
A;Cross-references: EMBL:AB034704; PIDN:BAA94063.1
A;Experimental source: strain III44
C;Genetics:
C;Superfamily: phytoene dehydrogenase

Query Match 24.0%; Score 623; DB 2; Length 511;
Best Local Similarity 31.1%; Pred No. 1.7e-40;
Matches 154; Conservative 106; Mismatches 215; Indels 20; Gaps 10;

10 IVIGAGLGSLSAISLATAGFSVQLIEKNDKVGKLNIMTKDGTFDILGSPILTMPIFE 69
22 LVVSGFGGMAAAVRLAAGYRVTVLEKLDAPGGRAYVHREGHVFDAGPTIVTVPLFD 81
70 ALFTGAGKMMADYVQIQVBPWHNFEDSGVIDLCEAETORRELDKLGPQTYAQFQ 129
82 ELWALAGRKFSDIIEKSLDPFYRIKFDGHPFYSYGDPAARMRAEVRIRISDAEGFERF 141
130 LDYSKNLCTETETAGYFA---KGLDGFMDLLKFKYGPLRSLLSFDVFRSMOQGVRRFISDPK 186
142 M-READQC--YELGPRTLGDKAFDTVGDLIK-AAFL--IVKLRGWRSLHQAQVSSHLKPK 195
187 LVEILNFIKYVGSSPYDAPALMNLPPYQVHYGLWTVKGMYGMAQAKLAVELGVEI 246
196 LRIASLOSLLIGGNPFSVTSYALVNALERQMGVHWAMGTGELIRGLVDVFGMGGTM 255
247 RLDAEVSIIQKQDGRACAVKLANGDVLDPADIVVSNMVEVIPAMEKL--RSPASELKMM 302
256 RLKAEVKRIEVDNGVATGVTLADGERIPADIVVNCNGDTGYLYKNLVDARMKHWT-DAR 314
303 QRFEPSCGLVHLGVDRLYPOLAHNFFSYDHPREHFDVFKSHRLSDPTIYVAPCK 362
315 ERGHYSMGLFWYVFGTDRRYEDVPHHMMVLPGRRELDDIFRKKCLASDPSIFLHRTA 374
363 TDPAQAPAGCEIILKILPHIL--DPDKLKTAEADYSALRERVLVKLERMGLTDLRQHIWT 420

b 375 TDPSMAPAGCDTFYALMPVPHLGGTDWTTQAEPY---RQSVQEALETRVLPGLGHLRV 431
y 421 EBYWTHLDIOAKYYSNCGSIYGVVADRFXNLGFKAPORSSELSNLYFVGGSVNPGSGWPM 480
b 432 SFCTTFLDFQHRLLSYKGAGFGLLEPLLOSAYFRPHNRSEDEVKNLFMWGASTHPGAGVPG 491
y 481 VTLSGQLVRDKIVAD 495
b 492 VMSAKAL-ESVLPD 505

Search completed: February 29, 2004, 14:52:41
Job time : 15.8338 secs